

AD-A202 912

DTIC FILE COPY

(4)

AFGL-TR-88-0279

ENHANCEMENTS TO THE AFGL STATISTICAL ANALYSIS PROGRAM (ASAP) FOR THE GLOBAL  
MULTIVARIATE ANALYSIS OF HEIGHTS AND WINDS

Ross N. Hoffman, Mark Mickelson and Thomas Nehr Korn

Atmospheric and Environmental Research, Inc.  
840 Memorial Drive  
Cambridge, MA 02139-3794

17 September 1988

Scientific Report No. 4

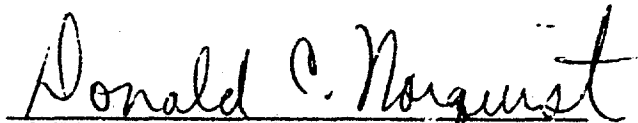
Approved for public release; distribution unlimited


AIR FORCE GEOPHYSICS LABORATORY  
AIR FORCE SYSTEMS COMMAND  
UNITED STATES AIR FORCE  
HANSCOM AFB, MASSACHUSETTS 01731-5000

DTIC  
ELECTE  
DEC 19 1988  
S H D

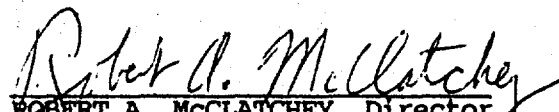
88 12 19 012

"This technical report has been reviewed and is approved for publication"

  
DONALD C. NORQUIST  
Contract Manager

  
DONALD A. CHISHOLM, Chief  
Atmospheric Prediction Branch

FOR THE COMMANDER

  
ROBERT A. MCCLATCHEY, Director  
Atmospheric Sciences Division

This document has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFGL/DAA, Hanscom AFB, MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

# REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFGL-TR-88-0279	
6a. NAME OF PERFORMING ORGANIZATION Atmospheric and Environmental Research, Inc.	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Geophysics Laboratory	
6c. ADDRESS (City, State, and ZIP Code) 840 Memorial Drive Cambridge, MA 02139		7b. ADDRESS (City, State, and ZIP Code) Hanscom AFB, MA 01731-5000	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-86-C-0141	
8c. ADDRESS (City, State, and ZIP Code)		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 6670
		TASK NO. 10	WORK UNIT ACCESSION NO. CB
11. TITLE (Include Security Classification) Enhancements to the AFGL Statistical Analysis Program (ASAP) for the Global Multivariate Analysis of Heights and Winds			
12. PERSONAL AUTHOR(S) Ross, N. Hoffman, Mark Mickelson, Thomas Nehrkorn			
13a. TYPE OF REPORT Scientific Report No. 4	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) 1988 September 17	15. PAGE COUNT 110
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This report documents changes to and experiments designed to test the impact of these changes on the AFGL Statistical Analysis Program (ASAP) of heights and winds on the model sigma surfaces. Key enhancements include a new, simpler and more efficient method of converting pressure layer temperatures into sigma level heights, a corrected and improved handling of the least squares equations and a new solver for these systems of equations based on standard LINPACK routines. We conducted several offline tests of the new versions of procedures SATLTMP and ASAP2 which embody these enhancements to verify their correctness, to compare their results to results of the earlier versions and to tune certain control constants. We then conducted a one week data assimilation experiment, using simulated data to examine the impact of the enhanced ASAP. (continued on reverse)			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL D. Norquist		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL AFGL/LYP

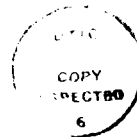
19. (Continued)

The changes made to ASAP have enhanced its consistency and operation. Separate studies of the satellite profile data preprocessing and the specification and solution of the normal equations demonstrated that the changes made are desirable. The overall positive impact of the enhancements is small as measured by a one week data assimilation experiment based on simulated data. *Key points*

+ 1 1 4

# TABLE OF CONTENTS

	<u>Page</u>
1. Introduction .....	1
2. Enhancements .....	1
2.1 Satellite temperature profile preprocessing .....	1
2.2 Specifying the normal equations .....	4
2.3 Solving the normal equations .....	6
2.4 Other minor issues .....	8
3. Tests .....	10
3.1 Testing SATLTMP in GRDSTA .....	10
3.2 Testing new CHLSKY and ASAP2 .....	11
3.2.1 Analysis of recalculation criteria .....	11
3.2.2 Tuning of new version of CHLSKY .....	17
3.2.3 Estimated analysis errors and corrections: impact and sensitivity .....	21
3.3 Impact test .....	24
4. Conclusions .....	30
5. Acknowledgements .....	30
6. References .....	31
Appendix A: Program Documentation .....	A-1



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

## 1. INTRODUCTION

The AFGL Statistical Analysis Program (ASAP) described by Norquist (1986, 1988) provides an optimal Gandin type analysis of height and winds on the sigma levels/layers of the AFGL Global Spectral Model (GSM). ASAP follows Dey and Morone (1985) and Bergman (1978, 1979) fairly closely. Detailed documentation of ASAP are given in an earlier report (Hoffman, 1987).

The present report documents changes to ASAP and experiments designed to test the impact of these changes. During preparation of Hoffman (1987), we noted some possible enhancements and minor corrections which should be made to ASAP. In the past year, a number of important enhancements have been made and tested. The current report documents these activities. Similar changes have been made to the moisture analysis program.

## 2. ENHANCEMENTS

Key enhancements include a new, simpler and more efficient method of converting pressure layer temperatures into sigma level heights, a corrected and improved handling of the least squares equations and a new solver for these systems of equations based on standard LINPACK routines.

### 2.1 Satellite temperature profile preprocessing

Satellite temperature profile data is generally reported as layer mean temperatures or equivalent height differences between mandatory pressure levels. These data must be anchored to the forecast height field (as is done in ASAP) or to a preexisting surface pressure or 1000 mb height analysis. Additionally, these data must be interpolated to the sigma levels of the model. Although this vertical interpolation might be accomplished statistically, ASAP interpolates all profile data analytically in the vertical. For satellite temperature profile data we now accomplish this transformation by converting the mean pressure layer temperature ( $T_\ell$ ) to an equivalent temperature profile defined by the values of pressure level temperatures ( $T_\ell^\wedge$ ) making use of the assumption that temperature is piecewise linear in  $\ln p$ .

This temperature profile is then integrated hydrostatically to obtain sigma layer temperatures ( $T_k$ ) and the sigma level heights ( $Z_k^\wedge$ ). In principle, this transformation is not necessary, the analysis scheme could make use of the fact that  $Z_k^\wedge$  is correlated with  $T_\ell$ . In practice, it is desirable to make this transformation, since the correlations with individual  $T_\ell$  are weak.

By consistently making the assumption that  $T$  is linear in  $\ln p$  between the  $p_\ell^\wedge$ , we will introduce no height errors due to our transformations for any layer which has been integrated over. We note that in converting layer to level temperatures, there is an extra degree of freedom in the level representation. We eliminate this by assuming that our piecewise linear temperature profile has the same slope in the two layers closest to 400 mb ( $p_a$ ).

The subscript notation used here is the following:  $\ell$  is reserved for pressure,  $k$  for sigma. A caret over the first character of a subscript indicates a level quantity. Subscripts increase with increasing height (decreasing pressure). Layer  $\ell$  is bounded by levels  $\ell$  and  $\ell-1$ . The first level has an index of zero. We use  $q$  to denote  $\ln p$  throughout.

We assume here that the satellite temperatures are actually thicknesses converted to temperatures hydrostatically. That is, they are temperatures averaged with respect to  $\ln p$  (L. McMillin, pers. comm., 12 Feb. 88). Since  $T$  is linear on  $q$ ,

$$T_\ell = (T_\ell^\wedge + T_{\ell-1}^\wedge)/2$$

Therefore, given the  $T_\ell$  and one  $T_\ell^\wedge$ , we can then step up

$$T_{\ell+1}^\wedge = 2T_{\ell+1} - T_\ell^\wedge$$

and down

$$T_{\ell-1}^\wedge = 2T_\ell - T_\ell^\wedge$$

to determine all the  $T_\ell^\wedge$ . If  $p_k^\wedge$  lies in layer  $\ell$ , i.e. if  $p_{\ell-1}^\wedge \geq p_k^\wedge \geq p_\ell^\wedge$ , then we may calculate  $T_k^\wedge$  by interpolating in  $q$ :

$$T_k^\wedge = T_\ell^\wedge + [(T_{\ell-1}^\wedge - T_\ell^\wedge)/(q_{\ell-1}^\wedge - q_\ell^\wedge)](q_k^\wedge - q_\ell^\wedge)$$

To determine the first  $T_{\ell}^{\wedge}$ , we find the  $m$  such that  $p_m$  minimizes  $|p_m - 400|$  and assume that  $T = a + bq$  in layers  $m$  and  $m + 1$ . Thus,  $T_m$  is the temperature at  $(q_m^{\wedge} + q_{m-1}^{\wedge})/2$  and  $T_{m+1}$  is the temperature at  $(q_{m+1}^{\wedge} + q_m^{\wedge})/2$ . As above, we interpolate in  $q$  to obtain

$$T_m^{\wedge} = T_m + [(T_{m+1} - T_m)/(q_{m+1}^{\wedge} - q_{m-1}^{\wedge})](q_m^{\wedge} - q_{m-1}^{\wedge}).$$

Given the  $T_{\ell}^{\wedge}$  and  $T_k^{\wedge}$ , we calculate the sigma layer mean temperatures by dividing the sigma layer by the pressure levels. We calculate  $T_k$  only if the  $k$ th sigma layer is within our temperature profile, i.e. if  $p_0^{\wedge} \geq p_{k-1}^{\wedge}$  and  $p_k^{\wedge} \geq p_L^{\wedge}$ . In general, sigma layer  $k$  will be covered by pressure layers  $\ell=m, n$ . The contribution of layers  $m$  and  $n$  may be due to only part of the layer. In all cases, the contribution to  $\int T dq$  from any one layer (or partial layer) will be the thickness of the layer times the average temperature of the layer which is the average of the temperature at the two end points. If  $B$  and  $T$  indicate bottom and top, then

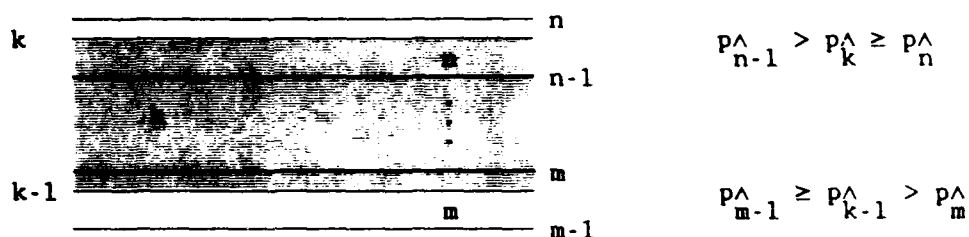
$$T_k(q_{k-1}^{\wedge} - q_k^{\wedge}) = \sum_{\ell=m}^n (q_B - q_T)(T_B + T_T)/2$$

where

$$q_B = \begin{cases} q_{k-1}^{\wedge} & \text{if } \ell = m \\ q_{\ell-1}^{\wedge} & \text{otherwise.} \end{cases} \quad q_T = \begin{cases} q_k^{\wedge} & \text{if } \ell = n \\ q_{\ell}^{\wedge} & \text{otherwise.} \end{cases}$$

$$T_B = \begin{cases} T_{k-1}^{\wedge} & \text{if } \ell = m \\ T_{\ell-1}^{\wedge} & \text{otherwise.} \end{cases} \quad T_T = \begin{cases} T_k^{\wedge} & \text{if } \ell = n \\ T_{\ell}^{\wedge} & \text{otherwise.} \end{cases}$$

Note that  $m$  might be equal to  $n$  if the sigma layer is within the pressure layer. In general we have:



Once we have the  $T_k$  values, we obtain the heights hydrostatically by

$$Z_k^{\wedge} = Z_a + (R/g) T_k (q_{k-1}^{\wedge} - q_k^{\wedge})$$

if  $T_k$  is not missing, and  $Z_k^{\wedge}$  = missing if  $T_k$  is missing. The anchor for level  $k$  is

$$Z_a = \begin{cases} Z_* & \text{if } k = 1 \\ Z_{k-1}^{\wedge} & \text{if } Z_{k-1}^{\wedge} \text{ is not missing} \\ \hat{Z}_k^{\wedge} & \text{otherwise} \end{cases}$$

where  $\hat{Z}_k^{\wedge}$  is the first guess value.

We have implemented these suggestions in subroutine ANCHOR which is called by a new version of subroutine SATLTMP which directly replaces the old version. Comparisons of the old and new versions (2.0 and 3.0) of SATLTMP are reported below (Section 3.1) and the ANCHOR and SATLTMP (version 3.0) procedures are described in detail in the Appendix.

## 2.2 Specifying the normal equations

Three inconsistencies in specifying the normal equations have been corrected.

- 1) The geostrophic factor acts to decouple the wind and height analysis in the equatorial belt. A strict decoupling is now enforced in this belt. If desired, it would now be simple to decouple the wind components from each other to analyze a divergent wind field. As of September, 1984, the NMC tropical wind analysis is univariate in the equatorial belt (Dey et al., 1985a, Section D.2).

Previously, intercorrelations poleward of  $10^\circ$  resulted in analyses which were not strictly univariate in  $Z$  and bivariate in  $(u,v)$ . To force decoupling some of the analysis weights were set to zero (see ASAP2(40) in Hoffman, 1987). Consequently the estimate of  $A_E$  produced was not strictly correct, since the weights used are no longer the least squares solution. That is, in this case Bergman's Eq. 2.8 and 2.11 do not hold and his Eq. 2.6 should be used to estimate  $A_E$  exactly.

For analysis grid points equatorward of  $10^\circ$  we now force a strict, proper univariate  $Z$  analysis and bivariate  $(u,v)$  analysis by setting all  $u_z$ ,  $v_z$ ,  $z_u$  and  $z_v$  correlations to zero in the observation-observation matrix and the observation-grid-point RHS (right hand sides). This is equivalent to solving two separate systems of equations since the matrix is now a permutation of a block diagonal matrix. The approach we have taken minimizes changes to the code but involves some extra computations (with zeros) since factorizing a matrix goes like  $N^3$ . Since the RHS have zeros for the crossterms we are assured that the corresponding analysis weights will be zero. However we have the convenience of retaining the preexisting data structure since all the data are kept in the update. We have actually set the criterion for equatorial belt as  $11^\circ$  so that the Gaussian grid points at  $10.08^\circ$  are included.

- 2) The calculation of the RHS has been corrected. In the selection phase the RHS are normalized by  $(1 + \epsilon^2)$ , i.e., by the diagonal of the correlation-correlation matrix. The normalized RHS are the analysis weights we would obtain if each observation were the only one. However this normalization is not now applied to the actual RHS of the normal equations.
- 3) An underground analysis is performed so that the layer  $Z$  analysis may be interpolated to level  $T$  increments which are then converted to layer  $T$  increments by the Flattery algorithm.

Although only a  $Z$  underground analysis is required we now perform a complete  $(Z,u,v)$  underground analysis. We want the lowest two layer and

the underground analyses to be as similar as possible to assure reasonable agreement between the surface and first level temperatures. Therefore our procedures are exactly the same for the underground layer as for the other layers. This includes data selection and the solution of the normal equations, including data rejection as necessary (described below), even if only the u or v analysis is not reasonable.

Previously, the underground analysis was handled specially: Potentially, the underground Z analysis could have been very different from the lowest layer analysis because of differences in data selection.

### 2.3 Solving the normal equations

The least squares analysis or normal equations are solved at each grid point for (Z,u,v). It is desirable that precisely the same data are used for each of Z, u and v so that the increments are geostrophically balanced.

Previously, the Choleski algorithm as implemented by Stobie (1984; Stobie et al., 1985) was used to solve the normal equations. The version of this algorithm we had been using does not properly check for near singular or non-positive definite systems. Consequently, some of the analysis weights  $a'$ , were not properly determined. In some cases these incorrect weights may be quite large in magnitude and lead to impossible values of the estimated analysis error  $A_E$ . These conditions were checked for and if necessary an observation was deleted and the normal equations resolved. The criteria for deleting an observation are discussed at greater length in Section 3.2.1. Since each of Z, u and v were solved for independently, different data might be used for each. Worse, improper, but reasonable weights were used in some cases.

The solution of the normal equations is now implemented as decomposition into  $R^T R$  form followed by a forward-backward solution for each RHS. We used the LINPACK routines SPOCO and SPOSL (Dongarra et al., 1979) to factor and solve the system. In addition the LINPACK routines are used to calculate RCOND, an estimate of the reciprocal of the condition number of the matrix. Calculating RCOND involves some extra calculations. If the matrix is of order

N then SPOCO is reported to take  $(1 + N/18)$  times the time to simply factor the matrix.

If RCOND is too small, say less than 0.001, then we eliminate an observation and recalculate the solution. This is similar to what was done before, but uses a more rational criterion. In choosing the cutoff for RCOND we have the following guidance: If we are solving  $Ax = b$  and if the reciprocal condition number is equal to  $10^{-d}$  then we expect the solution  $x$  we calculate to have  $d$  fewer significant digits than the elements of  $A$ . While we calculate the elements of  $A$  to full machine precision this calculation is based on simple models of the statistical functions. It is difficult to argue that we really know these correlations to much more than 2 decimal places. If reciprocal condition number is 0.001 in a particular case then we have no actual knowledge of the proper weights to use.

Of course the value RCOND from SPOCO is an estimate. To be safe we retain loose checks on  $a'$  and a check on  $A_E$ . To avoid updating the pointers and creating secondary arrays we leave a placeholder equation for each observation deleted. For the  $i$ th observation this equation is  $a_i = 0$ . That is we set all correlations with the deleted observation to zero and leave a one on the diagonal. Obviously, we will obtain a zero analysis weight for this observation.

As a possible alternative we note that the factorizations produced by the LINPACK routines allow one to solve any leading subproblem. That is, the factorizations could be used to solve the system obtained by dropping the last one or several, rows and columns. If data were presorted so that the first datum to be dropped (if it proved necessary) were that last datum in the normal equations and the second to be dropped were the next to last datum, etc., then only a single decomposition would be required. However this presorting would likely be more expensive than recalculating just a small percentage of solutions as is done now.

## 2.4 Other minor issues

A number of other minor issues have been dealt with. These include:

- 1) The translation of Quality Indicators (QI) from FGGE Level II data to internal ASAP Quality Marks (Q) has been corrected to properly account for TWOS, DROPWINDSONDE and COLEBA data. Subroutine CQCV performs this translation. Subroutine CQCV has been rewritten as described in the appendix. Here we describe our interpretation of Table IV of Appendix A of WMO (1986, pp. 5-7) which explains the QI for type 1 and type 2 data.

QI is composed of two digits denoted IH and IV, nominally for horizontal and vertical checks. For radiosondes, pilot balloons and aircraft data this translation is discussed by Hoffman (1987, pp. 81-82). For TWOS, DROPWINDSONDE and COLEBA data the IH digit indicates instrument uncertainty 1 being very low and 9 being very high with 0 indicating unknown, and the IV digit indicates a measurement type with 1 being very precise and larger numbers indicating a larger averaging or interpolation interval. Again 0 indicates unknown.

We adopt the following translation (see Table 1):

- 1) For TWOS and DROPWINDSONDE we flag data as suspect if
  - IH > 5 (uncertainty in V > 8 m/s; in Z > 300 m; in T > 3.2 degrees; in RH > 16 %)
  - or IV > 4 (a vertical interval of more than 1.1 km).
- 2) For COLEBAs we flag data as suspect if
  - IH > 5 (uncertainty in V > 8 m/s)
  - or IV > 1 (not a standard Doppler solution)
- 3) For both, an IV = 0 flag is ignored and the quality mark is set on the basis of IH. Otherwise we translate IH and IV separately and take the maximum as the quality mark. (Note (\*)ed-items in translation tables account for the special cases.)

NOTE: We have lumped idsi = 13 and 14 as TWOS data, although WMO (1986) is not precise about idsi = 13 quality indicators.

Table 1. Translation of quality indicators (QI) into quality marks (QM)

QI	0	1	2	3	4	5	6	7	8	9	ids1
IH QM	*0	0	6	8	2	5	3	4	7	99	11, 12, 20-29
	1	0	0	0	0	0	6	6	6	6	13, 14, 15
	1	0	0	0	0	0	6	6	6	6	16
	1	1	1	1	1	1	1	1	1	99	all others
IV QM	1	0	6	8	2	5	3	4	7	99	11, 12, 20-29
	*0	0	0	0	0	6	6	6	6	6	13, 14, 15
	*0	0	6	6	6	6	6	6	6	6	16
	1	1	1	1	1	1	1	1	1	99	all others

- 2) U.S.A. Special Effort quality controlled and/or corrected satellite temperature profiles are now preferred to raw reports. Previously, the selection was the other way around, effectively deleting all special effort reports. Tables 37 and 38 of Appendix A of WMO (1986) are used. See MASTOR4 in the appendix for details.
- 3) U.S.A. Special Effort quality controlled cloud drift winds are now preferred to raw reports. Previously, the selection was the other way around. Table 36 of Appendix A of WMO (1986) includes the codings of the USA Special Effort. The translation used and other details are described in MASTOR6 in the appendix.
- 4) The calculation of uu and vv forecast error correlation used in the buddy check criterion has been corrected. In FLAGS(19), these quantities are now calculated without taking their absolute values.
- 5) Radiosonde dewpoint temperatures are now converted to RH in MASTOR1(8) and then interpolated linearly in ln p in PTOSIG(20).
- 6) Data rejected in FLAGS by the gross and buddy checks are reported in a new file on unit 9 by REJECT.

### 3. TESTS

We conducted several offline tests of the new versions of procedures SATLTMP and ASAP2 to verify their correctness, compare their results to results of the earlier versions and to tune certain control constants. We then conducted a one week data assimilation experiment, using simulated data to examine the impact of the enhanced ASAP.

#### 3.1 Testing SATLTMP in GRDSTA

In order to test the new version of SATLTMP (version 3.0), we compared both it and the old version (version 2.0) to procedure PTOSIG. In the tests PTOSIG was presented with a radiosonde profile, that is height (Z) and temperature (T) at a series of mandatory (and possibly significant) pressure levels (p), to interpolate to Z values on the sigma level and T values in the sigma layers. Then each version of SATLTMP was presented with a profile of pressure layer mean temperatures derived hydrostatically from the radiosonde heights, to interpolate to sigma heights and temperatures. These latter values were then compared to the corresponding values obtained from PTOSIG (considered to be the truth here) and rmse and bias statistics accumulated. Before summarizing the results we briefly describe the three procedures and their underlying assumptions.

Procedure PTOSIG was developed at NMC and is described in detail in Gerlach (1983, pp. 44-46). PTOSIG assumes temperature (T) is a linear function of  $\ln p$  to interpolate heights from pressure levels to sigma levels. Sigma level heights are then differenced to give sigma layer temperatures. To interpolate height within any pressure layer PTOSIG uses the heights at the bounding interfaces and the temperature difference across the layer.

SATLTMP version 1.0 was based on the Flattery scheme as described by Norquist (1986, pp. 55-58). The Flattery scheme finds a least squares solution to two conflicting assumptions about the temperature profile, namely that layer temperature is the average of the bounding interface values and that the interface values may be obtained by interpolating the layer values linearly in  $\ln p$ .

SATLTMP version 2.0 is an alternative offered by Norquist (1986, pp. 64-65) in which layer temperatures are assumed to be mass weighted layer averaged temperatures and temperature within a layer is assumed to vary linearly with  $\ln p$ . However as noted above, satellite profile layer temperatures are really height differences expressed as (virtual) temperatures hydrostatically, i.e. averaged with respect to  $\ln p$ , not  $p$ . Version 2.0 still uses the Flattery scheme to get one sigma level temperature near 400 mb.

SATLTMP version 3.0, described above in Section 2.1, assumes that temperature varies linearly with  $\ln p$  within each pressure layer. It is also assumed that this relationship is differentiable across the pressure level closest to 400 mb.

In the following test cases both PTOSIG and SATLTMP were presented data defined at or between the mandatory pressure levels 1000, 850, 700, 500, 400, 300, 200, 100, 70, 50. These tests were conducted by making minor changes to the grid to station verification program of Norquist (pers. comm., 1987). Results of the test of SATLTMP versions 2.0 and 3.0 (SATONE and SATTWO) are compared here. The sample runs from 79 February 8 1200 GMT through February 15 0000 GMT excluding February 13 1200 GMT.

The numerical results are displayed in Tables 2 and 3 for height at sigma levels and temperature at sigma layers respectively. The results are also displayed graphically in Figs. 1 and 2. Clearly, except for the eleventh level/layer, both schemes are adequate and roughly comparable. However version 3.0 has smaller biases and rmse and is definitely better for the highest level/layer. Version 3.0 is also simpler in its assumptions and coding, and more efficient as judged by the computer timings of the test cases. Timings for a single synoptic time give 2.79 and 2.43 cp seconds of execute time for versions 2.0 and 3.0 respectively for a net savings of more than 10%. However version 3.0 does not process as many levels as version 2.0.

### 3.2 Testing new CHLSKY and ASAP2

#### 3.2.1 Analysis of recalculation criteria

To see the relationship between old and new criteria for rejecting data and recalculating the weights, we have extracted all data needed to rerun the

Table 2. Height differences statistics (m) comparing SATLTMP Versions 2.0 and 3.0 versus PTOSIG.

LEVEL	SIGMA	Version 2.0			Version 3.0		
		RMSE	BIAS	SAMPLE	RMSE	BIAS	SAMPLE
1	.925	2.9253	-.0917	2270	2.8628	.2145	2245
2	.800	9.4633	2.8972	4478	5.4872	-.2177	4411
3	.650	9.2876	2.1642	7152	5.7740	.6397	7015
4	.500	8.7859	.9099	7484	4.2194	-.0470	7351
5	.375	12.7985	3.0962	7827	10.2658	1.6389	7815
6	.300	14.9905	1.1486	7719	13.5419	1.9475	7677
7	.250	18.9387	-.8738	7753	17.2810	2.4887	7703
8	.200	23.6651	-6.7386	7185	20.2309	-1.6692	7139
9	.150	35.8403	-14.3255	7054	30.8529	-7.0962	7002
10	.100	46.9694	-5.9414	6161	16.7813	2.3607	6110
11	.050	123.2643	-43.2206	1574	8.8502	-.8772	1445
12	.010	-	-	0	-	-	0

Table 3. Temperature difference statistics (K) comparing SATLTMP Versions 2.0 and 3.0 versus PTOSIG.

LEVEL	SIGMA	Version 2.0			Version 3.0		
		RMSE	BIAS	SAMPLE	RMSE	BIAS	SAMPLE
1	.962	1.2810	-.0402	2270	1.2537	.0939	2245
2	.861	2.0373	.2547	4478	1.4436	-.0742	4411
3	.724	1.4712	-.0742	7152	1.1248	.1265	7015
4	.573	.6891	-.1486	7484	.5886	-.0897	7351
5	.436	.6426	.2511	7827	.7287	.1679	7815
6	.337	1.0466	-.2401	7719	1.1015	.0632	7677
7	.274	1.5945	-.2624	7753	1.6635	.0996	7703
8	.224	2.6612	-.9474	7185	2.3219	-.6406	7139
9	.174	2.7856	-.8259	7054	2.6414	-.6442	7002
10	.124	4.6059	.6537	6161	2.8510	.7667	6110
11	.730	6.1634	-1.6887	1574	.6359	-.0787	1445
12	.020	-	-	0	-	-	0

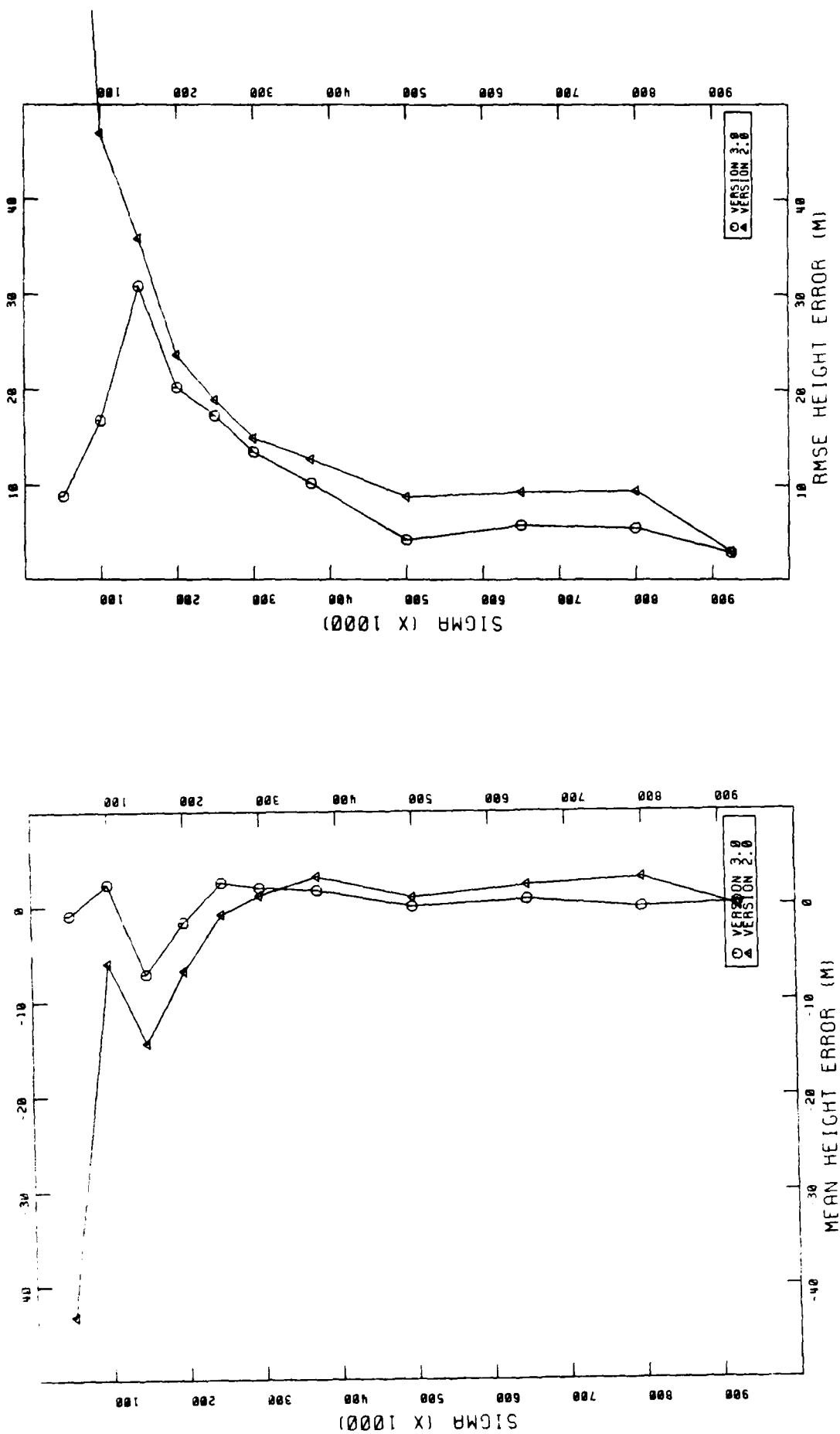


Fig. 1. Height differences statistics (m) comparing SATLTP Versions 2.0 and 3.0. Truth is taken to be radiosondes interpolated by PTOSIG. The sample consists of one week from SOP-1.

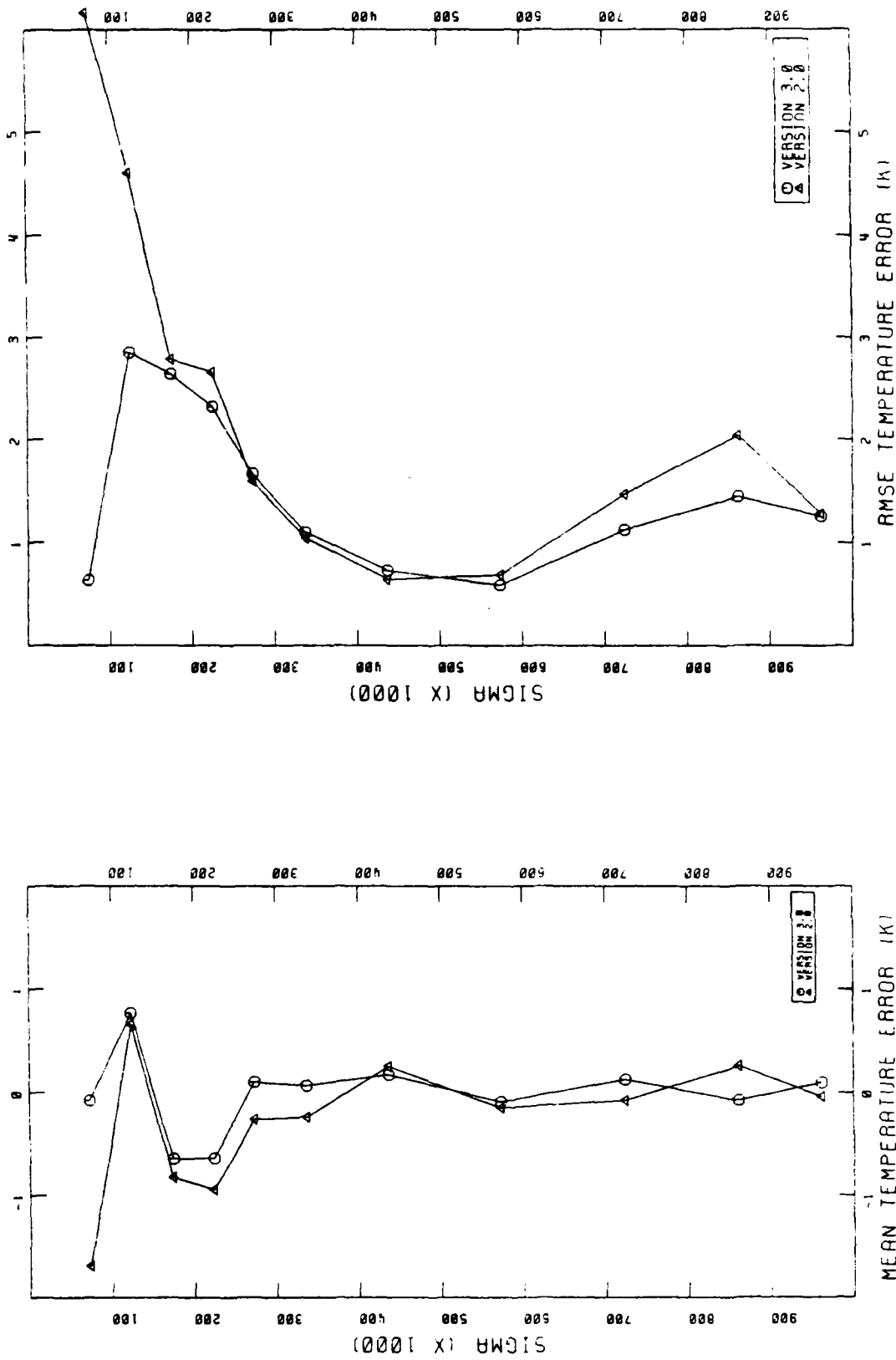


Fig. 2. Temperature differences statistics (K) comparing SATLTP Versions 2.0 and 3.0. As in Fig. 1.

least squares problem for a subset of the analyses for one synoptic time (06 GMT 17 June 1979) from our STATSAT experiment (Louis et al., 1987). The subset in question includes all systems of equations for a single level ( $K = 5$ ) which did not yield satisfactory solutions according to the old criteria.  $Z$ ,  $u$  and  $v$  problems were included. At level 5, out of a total of  $61 \times 62$  grid points there were 98 problem grid points. There were a total of 281 least square systems in our subset. For each problem grid point two or all three of the systems for  $Z$ ,  $u$  and  $v$  may be present and there may be multiple systems for each if it was necessary to remove more than one datum before a satisfactory solution was obtained.

For these cases, we have examined certain key quantities and their inter-relationships. These quantities are:

$A_E$         the normalized analysis error  
 PIVOT       the least pivot used by the CHLSKY subroutine  
 AP          the largest absolute value of the normalized weights  
 RCOND       the reciprocal of the condition number calculated by LINPACK  
              routine SPOCO. We set RCOND to 0 if the matrix is singular.  
 (Note: most of the plots identify RCOND as condition number.)

The original ASAP2 determines a solution is unsatisfactory if 1) CHLSKY finds a pivot whose absolute value is less than  $10^{-7}$ ; or 2) AP is larger than 1.1; or 3)  $A_E$  is not within  $[0,1]$ . In fact condition (1) never occurred in our sample, although there were many cases of negative PIVOT. Choleski pivots should never be zero or negative. This indicates a nonpositive definite matrix. Criterion (2) is arbitrary: some solutions with AP of 1.5 or 2.0 might be valid, while some solutions with small values of AP are actually poorly conditioned. Condition (3) only occurred when (2) had also occurred in the cases studied.

We note from the plots in Fig. 3 that large values of AP are associated with pivots of small absolute magnitude. Large values of AE are associated with small negative values of PIVOT and with large values of AP. We found that most of the cases of negative and small  $A_E$  (less than say 0.2) are associated with  $10.08^\circ\text{N}$ ,  $53.11^\circ\text{E}$ . These points are circled in some of the plots. We were unable to determine the difficulty with these systems.

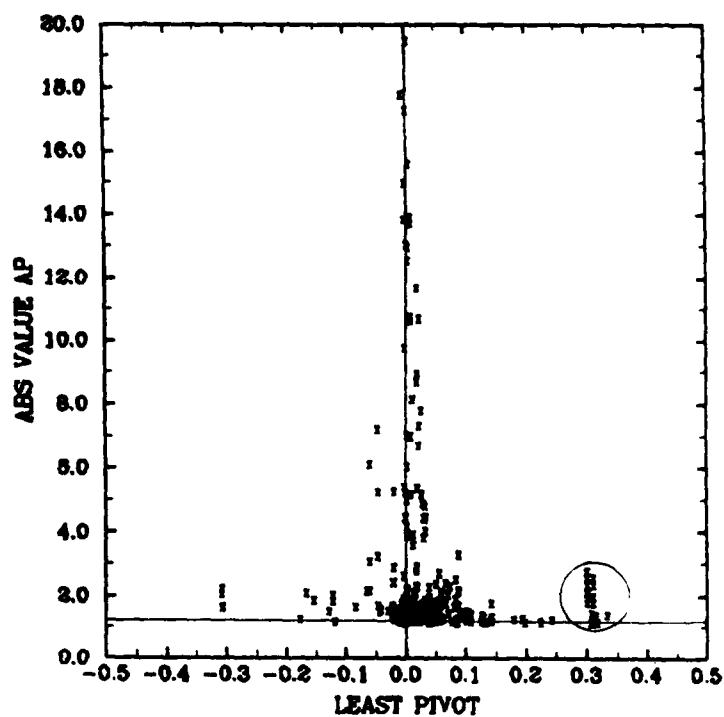
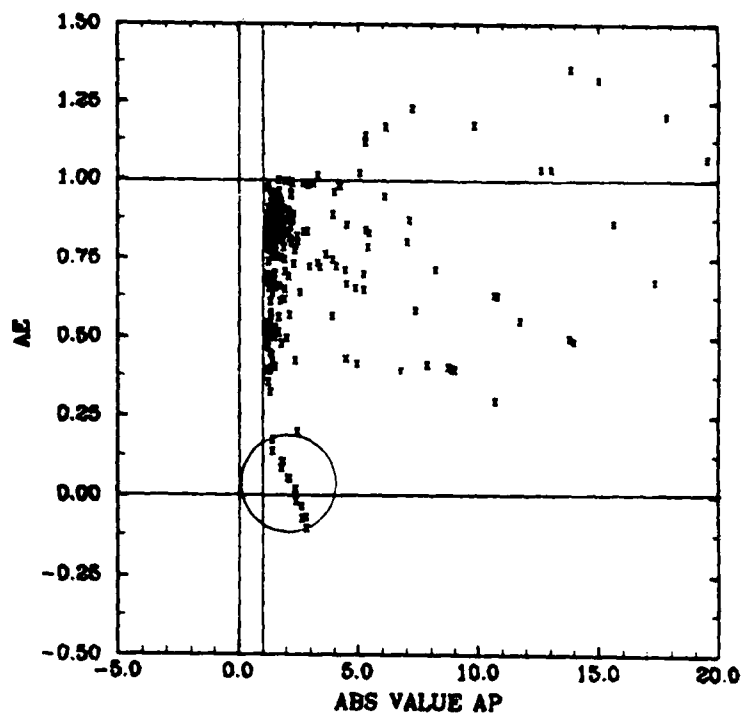
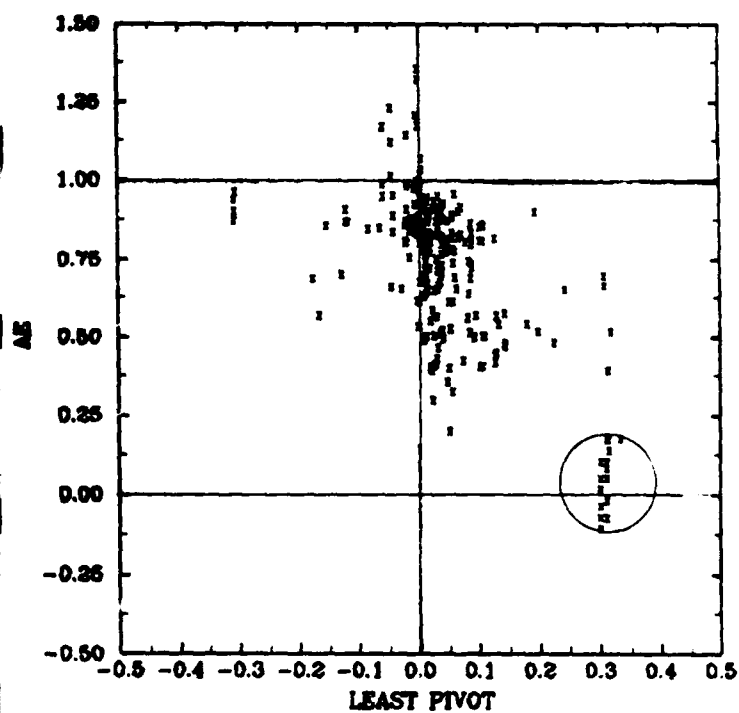


Fig. 3. Scatterplots describing relationships among parameters used by the original version of CHLSKY to check if the solution is satisfactory. See text for details.

We see from the plots in Fig. 4 that condition number is closely related to PIVOT and that large values of AP and  $A_E$  generally have small RCOND. Using a cutoff of 0.01 for RCOND would trap all but a few of the cases trapped by the old scheme in this sample.

We find that each time RECALC, the procedure to remove one of the most highly correlated data in the system, is invoked the condition number of the matrix improves, the AP maximum value decreases and the  $A_E$  value increases. The magnitude of these changes seems correlated. That is, if dropping a point significantly improves the condition number, it also makes a large difference to AP and AE.

Since the relationship between RCOND and AP is not well defined it seems preferable to eliminate data on the basis of RCOND which has a well defined mathematical meaning even if the cutoff value is chosen empirically. Fig. 5 shows that the estimate of reciprocal condition number from SPOCO compares well with the estimate from SCHDC, another LINPACK routine after we account for the fact that the estimates from SCHDC are generally overestimates, often by a factor of 10 (Dongarra et al., 1979, p. 8.14).

### 3.2.2 Tuning of new version of CHLSKY

The new version of CHLSKY has two adjustable parameters in the specification of the data rejection criteria. These are RCRIT and XCRIT. When either RCOND/RCRIT is less than one (i.e. the system is poorly conditioned) or XMAX, the largest normalized data weight, is greater than XCRIT, the most highly correlated datum is eliminated from the system and the reduced system is resolved. We conducted several test runs for the case of 06 GMT 17 June 1979 using the new versions of ASAP2 and CHLSKY as described in the appendix. The statistics reported below are based on all systems of equations for this case.

In the first run RCRIT = 0.001 and XCRIT = 4.0. In this run, except for extreme values of  $a'$ , all data rejection is based on the condition number of the system of equations. While there were many cases of  $A_E = 1$ , i.e. cases of no reduction in analysis error, there were only 4 cases of negative  $A_E$ . These impossible values of  $A_E$  were associated with marginal condition numbers (RCOND < 3/1000) and large AP values (XMAX > 3.0; three of these four cases had XMAX > 4.0). There were a total of 16 cases of XMAX > 4.0.

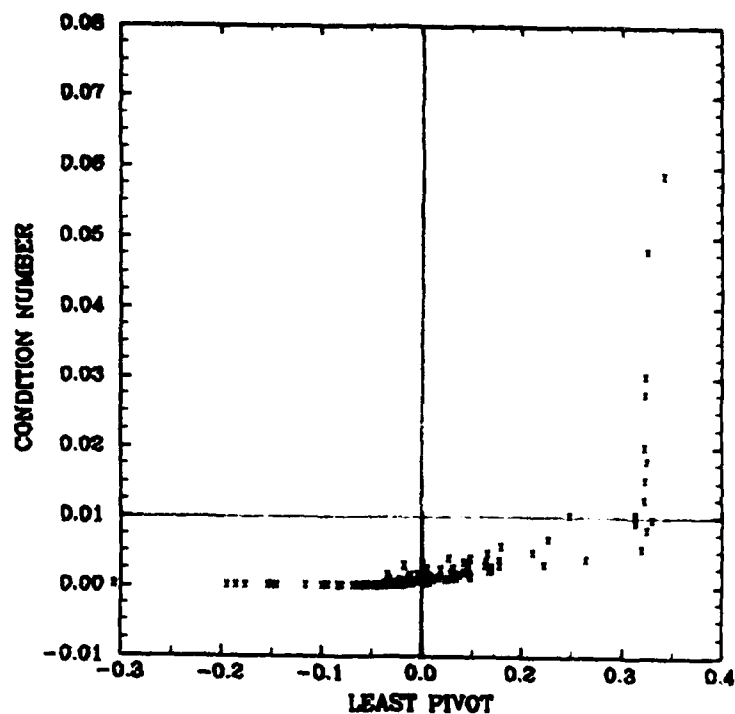
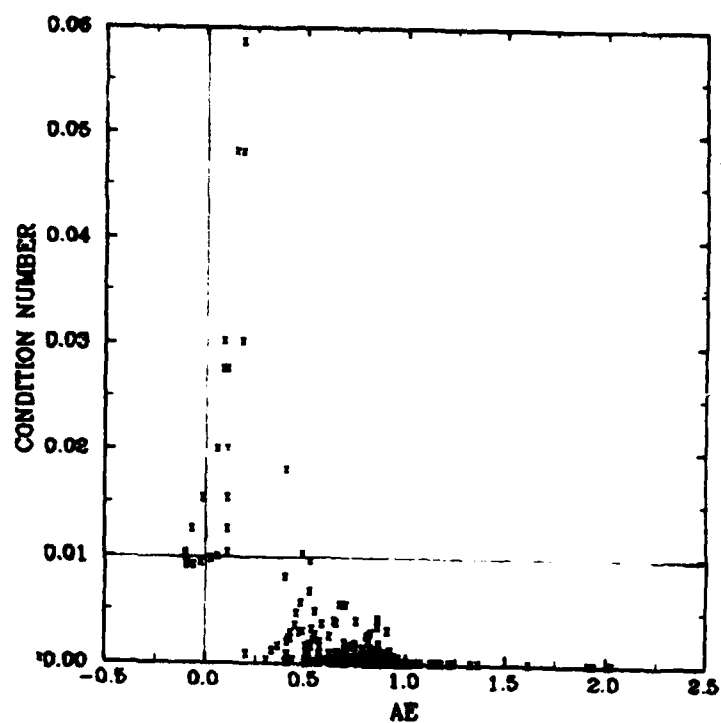
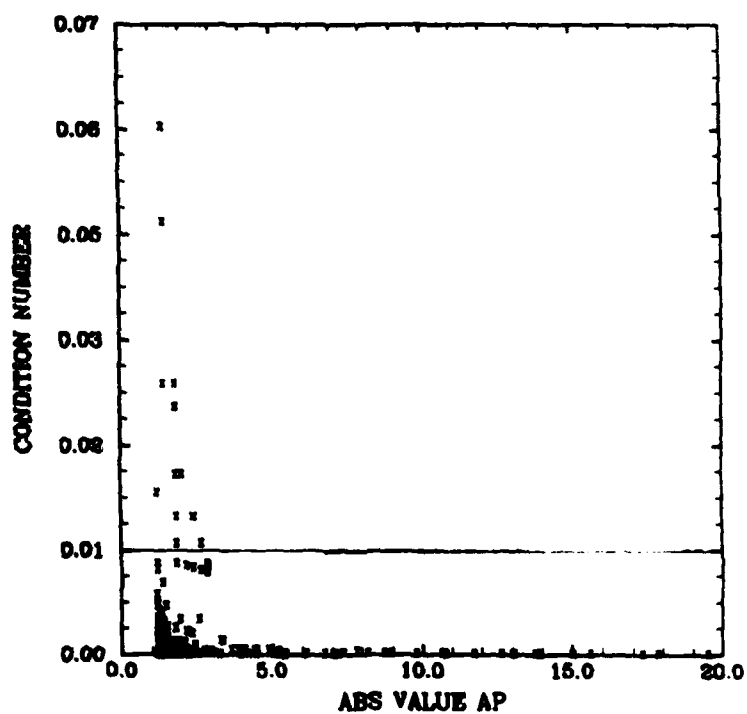


Fig. 4. Scatterplots describing relationships among parameters used by the original version of CHLSKY to check if the solution is satisfactory and reciprocal condition number calculated by LINPACK routine SPOCO. See text for details.

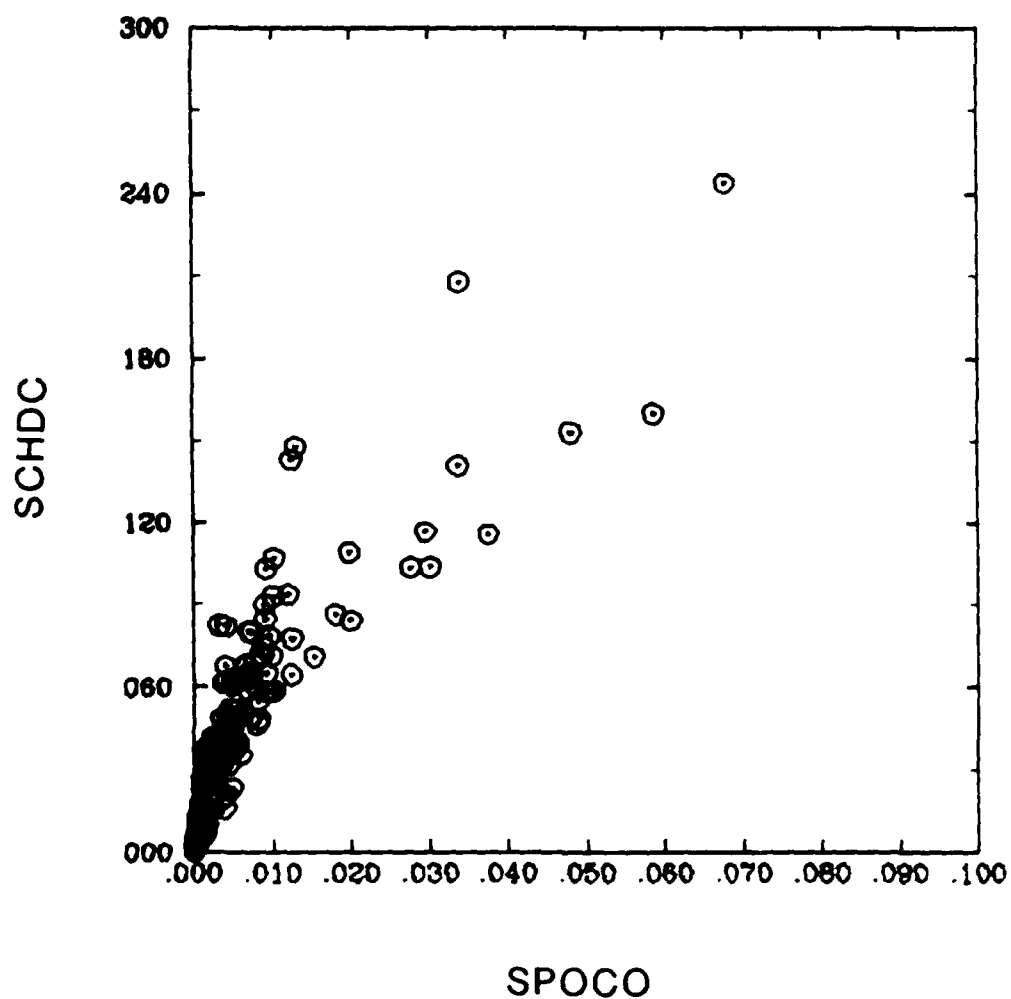


Fig. 5. Scatterplot describing relationship between reciprocal condition numbers calculated by LINPACK routines SPOCO and SCHDC. See text for details.

Out of a maximum possible 49166 grid points, roughly 37000 had some data. In total 14786 pieces of data at roughly 8100 grid points were eliminated in this run. Of these more than half, 8088 pieces of data at 4440 grid points, corresponded to non-positive definite systems, i.e. RCOND = 0.0.

We repeated the experiment outlined above with XCRIT = 1 and RCRIT = .00001. These values cause data rejection if the system of equations is non-positive definite but not if it is ill-conditioned unless a value of  $a' > 1$  is determined. By examining the distribution of XMAX for poor and well conditioned systems we see that many poorly conditioned systems have fairly reasonable values of XMAX. This is seen in Table 4, which is a table of histograms of the number of occurrences of different values of RCOND for different sub-ensembles defined by XMAX > 0, 1, 2, 3 or 4. The singular cases are not included here. The important result is that there are many poorly conditioned systems which have reasonable values of XMAX.

Table 4. Distribution in terms of reciprocal condition number, conditioned by size of maximum normalized weight.

1000*RCOND		Occurrences for XMAX >				
>	<=	0(*)	1	2	3	4
0.0	0.001	74	0	0	0	0
0.001	0.01	593	1	1	1	1
0.01	0.1	598	265	218	173	155
0.1	1	9270	1870	516	336	234
1	10	24274	2370	234	58	14
10	100	6402	408	0	0	0
100	1000	1370	1	0	0	0
0	1000	42581	4915	969	568	404

(\*) Estimate based on first 2000 systems of equations.

From this table we conclude that values of RCRIT of .001 and XCRIT of 2.0 are reasonable. With these values almost all data rejections will result from detecting poorly conditioned systems; only a few hundred additional data will be rejected by the other criteria. It does not seem worth while to put further effort into optimizing the rejection criteria for the following reasons:

- 1) The data being rejected are highly correlated with other data which is kept. It would be more worthwhile working on a methodology which combines the two data rather than rejecting one (e.g. Lorenc, 1981).
- 2) The majority of the data rejected are due to singular systems, not ill-conditioned ones. These data must be rejected in the current formulation of the OI.
- 3) A reformulation of the horizontal correlation function in terms of Bessel functions may result in better conditioned systems, reducing the importance of the rejection criteria.

### 3.2.3 Estimated analysis errors and corrections: impact and sensitivity

In order to assess the impact and sensitivity of the new choleski scheme on ASAP, we compared the estimated analysis errors and corrections from the old scheme (CONTROL), the new scheme (NEW) and one of the test cases (TEST) which was described above. NEW incorporates all the changes discussed in Section 2 (except for change 5 in Section 2.4) and makes use of RCRIT = 0.001 and XCRIT = 2. TEST incorporates only the changes discussed in Sections 2.2 and 2.3 and makes use of RCRIT = 0.001 and XCRIT = 4. We identify NEW - CONTROL as IMPACT, since it is a measure of the impact of all changes and NEW - TEST as SENSITIVITY since it is a measure of the sensitivity to small differences in data selection and in the parameters of the new choleski solver. Summary statistics are displayed in Tables 5, 6, 7 and 8 for height and u-wind component only. Results for v-wind component are similar to the u-wind component results.

Considering the mean EAE impact we see the new scheme results in estimates of the analysis error which are smaller by order one decameter in height and one m/s in wind components. The size of the variation of IMPACT is the same order of magnitude. By comparison EAE sensitivity is relatively small.

Table 5. Mean heights (m)

Level	EAE			Correction		
	NEW	IMPACT SENSITIVITY		NEW	IMPACT SENSITIVITY	
1	21.825	-1.045	.991	.025	-.192	-.405
2	24.689	-2.294	.107	-.364	-.724	-1.035
3	32.730	-3.667	-.085	-1.303	-1.597	-1.673
4	37.558	-3.932	-.042	-2.034	-1.867	-1.879
5	48.215	-4.696	-.058	-2.931	-2.584	-2.528
6	61.583	-6.258	-.057	-4.083	-3.470	-3.434
7	69.465	-7.641	-.069	-4.417	-3.552	-3.615
8	72.536	-8.431	-.088	-4.543	-3.529	-3.592
9	72.002	-8.631	-.122	-5.305	-3.572	-3.414
10	73.266	-9.135	-.147	-5.638	-3.480	-3.313
11	83.717	-6.313	-.074	-4.861	-2.312	-1.896
12	98.942	-1.248	.020	-1.323	.316	.769

Table 6. Standard deviation of height (m)

Level	EAE			Correction		
	NEW	IMPACT SENSITIVITY		NEW	IMPACT SENSITIVITY	
1	13.887	1.702	1.996	5.402	4.308	4.038
2	16.086	2.616	.789	6.810	4.322	3.556
3	19.709	4.340	.810	10.240	5.162	4.349
4	23.178	4.516	.892	11.931	4.708	3.866
5	28.119	5.244	1.098	15.307	5.398	4.527
6	34.450	7.089	1.407	19.633	7.203	5.910
7	36.519	8.479	1.635	22.919	8.308	6.415
8	35.119	9.026	1.743	24.735	8.832	6.688
9	31.917	9.322	1.832	27.512	9.902	7.831
10	26.856	9.403	2.102	31.446	12.126	9.237
11	18.541	6.405	1.063	28.169	11.151	10.221
12	11.184	2.246	.634	17.735	10.509	11.956

Table 7. Mean u-wind component (m/s)

Level	EAE			Correction		
	NEW	IMPACT	SENSITIVITY	NEW	IMPACT	SENSITIVITY
1	4.548	-.145	.058	-.008	.022	.021
2	5.197	-.248	-.005	-.027	.006	.019
3	6.982	-.372	-.008	-.040	.016	.022
4	7.938	-.405	-.002	-.050	-.004	.004
5	10.230	-.478	-.002	.009	.011	.019
6	13.072	-.627	-.002	-.020	.001	.017
7	14.611	-.806	-.007	-.008	-.001	.028
8	15.307	-.884	-.008	.027	.007	.023
9	15.532	-.835	-.014	.057	.040	.044
10	13.979	-.737	-.015	.114	.085	.072
11	15.042	-.559	-.023	.062	.084	.083
12	13.612	-.189	-.012	.127	-.003	-.034

Table 8. Standard deviation of u-wind component (m/s)

Level	EAE			Correction		
	NEW	IMPACT	SENSITIVITY	NEW	IMPACT	SENSITIVITY
1	2.288	.271	.184	1.170	.690	.569
2	2.695	.343	.154	1.312	.712	.549
3	3.769	.476	.166	1.664	.724	.572
4	4.167	.527	.171	1.783	.644	.497
5	5.336	.612	.206	2.159	.726	.598
6	7.008	.762	.261	2.785	.874	.652
7	7.750	.949	.299	3.279	.956	.684
8	7.843	1.007	.296	3.390	1.130	.876
9	7.596	.981	.283	3.249	1.102	.825
10	6.373	.905	.251	2.942	1.040	.781
11	6.462	.816	.330	2.549	1.052	.975
12	5.281	.579	.265	1.282	.746	.811

Note that the changes to SATLTMP effect the data values but not the observing error statistics. Therefore CONTROL and TEST corrections should be quite similar. This is especially evident in the mean height correction results where IMPACT and SENSITIVITY are nearly the same. The height corrections themselves are roughly half the size of typical radiosonde observing errors. The height IMPACT and SENSITIVITY are of order one decameter. Mean wind corrections are quite small; the size of the wind IMPACT and SENSITIVITY is of order one m/s.

We also examined horizontal maps of EAE for NEW, IMPACT and SENSITIVITY at .574 sigma. IMPACT is small, but persistent in the data rich areas and substantial and persistent in the oceanic data voids. SENSITIVITY is generally quite small and isolated. SENSITIVITY may be quite large in small areas in equatorial areas such as the Amazon Basin.

Actual impacts after spectral filtering and postprocessing are generally small (see Fig. 6). There are occasional noticeable small scale features. However changes will grow during the first few days of the assimilation cycle as the EAE stabilize. That is the new EAE will be used to forecast the EPE during the subsequent analysis. Note that the EAE is spectrally filtered during this process.

### 3.3 Impact test

We conducted a one week simulated data assimilation experiment with the old and new ASAP to study the impact of the enhancements described in Section 2. The control experiment, denoted STATSAT, was previously run as one of the baseline experiments in our studies of the impact of SSM/T-1 and SSM/T-2 and WINDSAT lidar wind data. The experiment using the new ASAP is denoted OITEST. (Change number 5 of section 2.4 was not implemented in OITEST, and the original version of the moisture OI was used in the assimilation.) The only differences between the two experiments are the enhancements to the ASAP described here. The nature run for these experiments is the ECMWF N48 grid point model simulation of 10 - 30 November 1979. The experiments begin at 0000 GMT on 18 November. To obtain realistic initial conditions we first ran a 96 h forecast from the nature run at 0000 GMT 11 November and then a 72 h spinup assimilation cycle. During the assimilations described here the full



FGGE Level II data set created by Dey et al. (1985b) exclusive of the lidar winds, was used in exactly the same fashion as described by Louis et al. (1987).

The experimental setup used is ideal for detecting impacts due to changes in the assimilation system. First of all, for verification purposes we know nature and can perform grid point to grid point comparisons. Second the data errors are ideal and known exactly. This eliminates the effect present in real data experiments of not knowing the errors statistics exactly. On the other hand these idealized errors are too easy to assimilate and results from such experiments are overly optimistic. As described by Dey et al. (1985b, Table 1), the simulated observational errors are all pure random numbers except that the satellite temperature errors include a small level by level bias depending on the total cloud amount as estimated from the relative humidity profiles. Also the nature run itself is smoother than the real atmosphere. Consequently, the analysis is able to average out the observational errors very effectively.

In general the results of the STATSAT and OITEST are very similar. It is difficult to determine which is better by examining synoptic fields. For example in the synoptic 1000 mb and 500 mb height fields at the end of the experiments at 0000 GMT 25 November it is evident that STATSAT and OITEST both agree very well with NATURE in large scale features. Smaller scale features tend to differ somewhat between STATSAT and OITEST, while NATURE has no smaller scales at all. As an example we show the 500 mb height fields and differences in the Southern Hemisphere in Fig. 7. The wind fields at 200 mb all look alike. Wind field errors are small and distributed fairly uniformly over the globe.

To quantify our results we calculated global mean and rms error statistics for STATSAT and OITEST. While OITEST is generally better than STATSAT, differences between the errors are small. In general the OITEST wind analyses show a small consistent improvement at all levels. The OITEST height analyses are slightly better at 1000 mb but slightly worse or no different from 850 mb through 150 mb. At 100 mb the OITEST height analysis is noticeably worse. At the uppermost two levels, 70 and 50 mb, the OITEST height and wind analyses are noticeably improved. Selected time evolutions of the global rms errors are shown in Fig. 8 and the time global rms errors are displayed in Fig. 9.

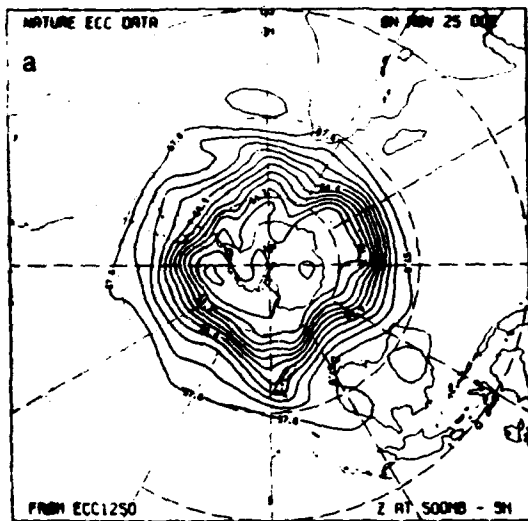
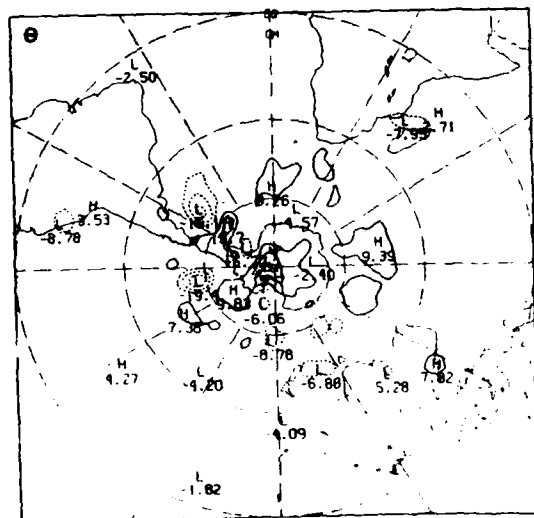
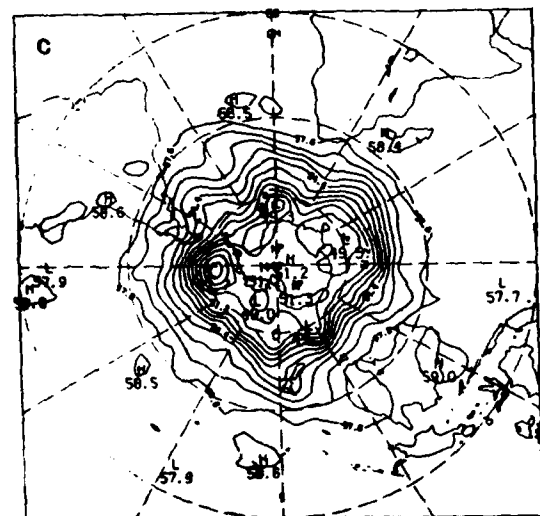
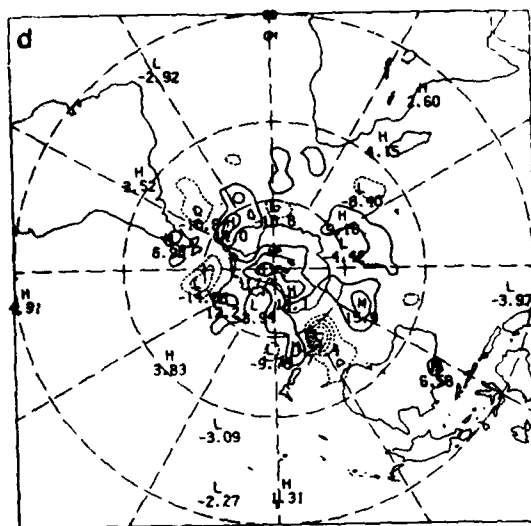
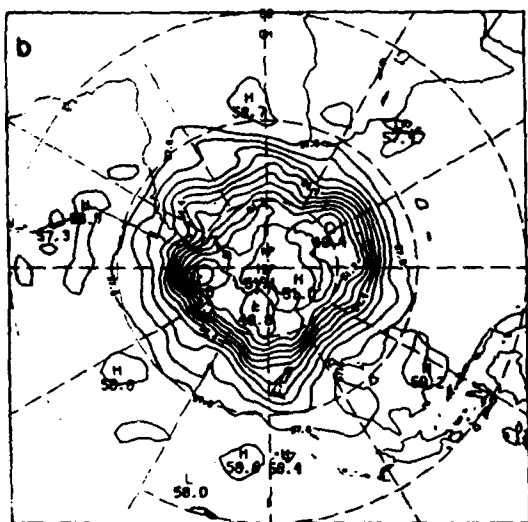


Fig. 7. Southern Hemisphere 500 mb height fields for 1979 November 17 0000 GMT: (a) NATURE, (b) STATSAT analysis, (c) OITEST analysis, (d) STATSAT error and (e) OITEST error.



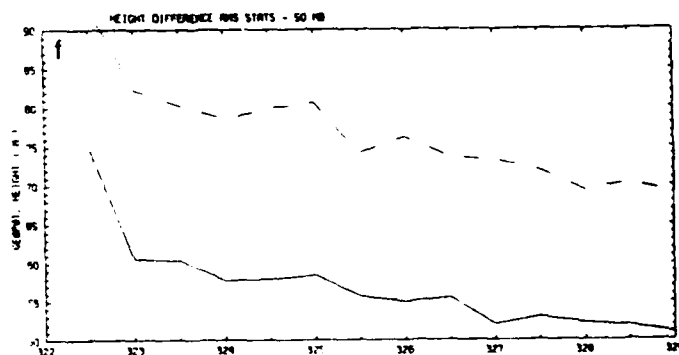
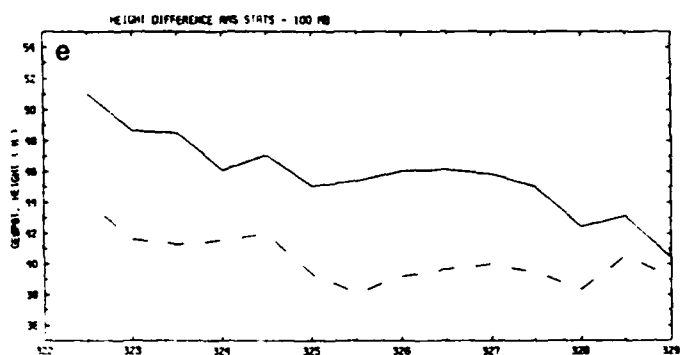
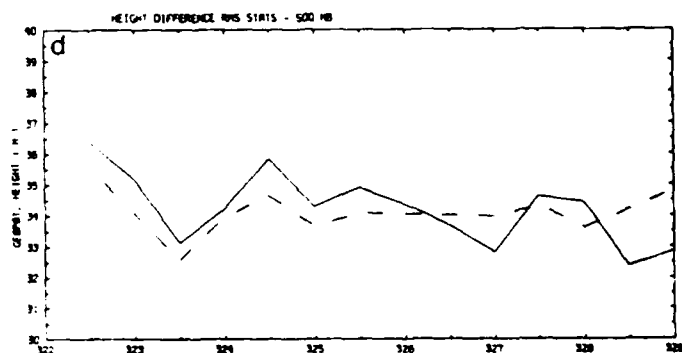
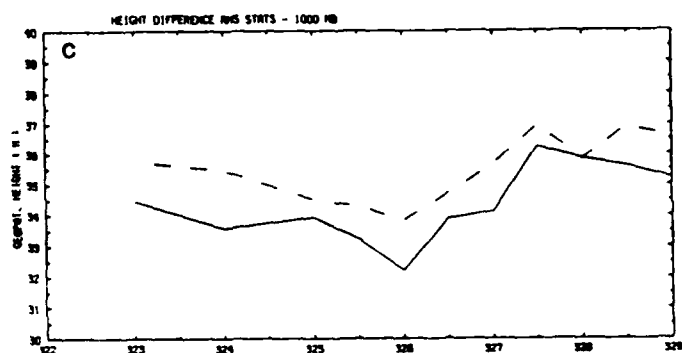
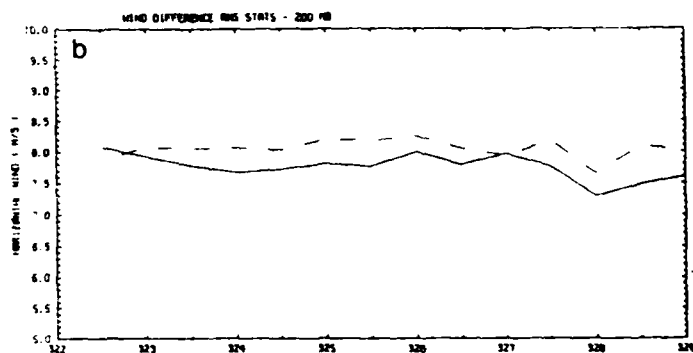
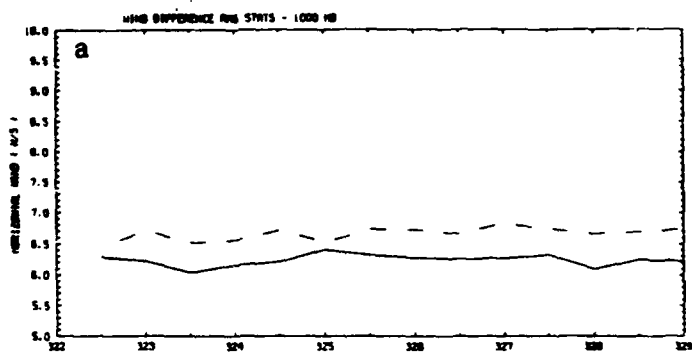


Fig. 8. Time evolution of global analysis error during the STATSAT (dashed) and OITEST (solid) assimilation experiments for rms vector wind error (m/s) at (a) 1000 mb, (b) 200 mb, and for rms height error (m) at (c) 1000 mb, (d) 500 mb, (e) 100 mb and (f) 50 mb.

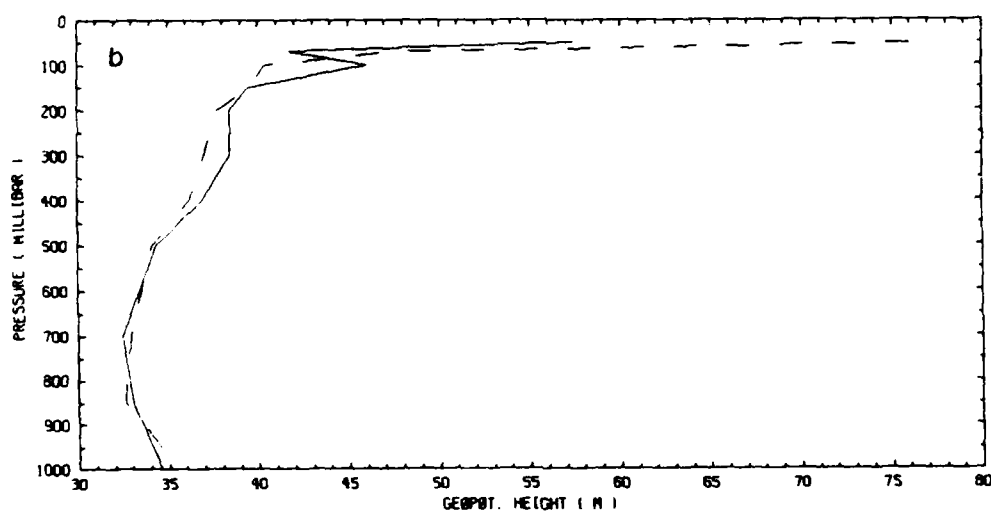
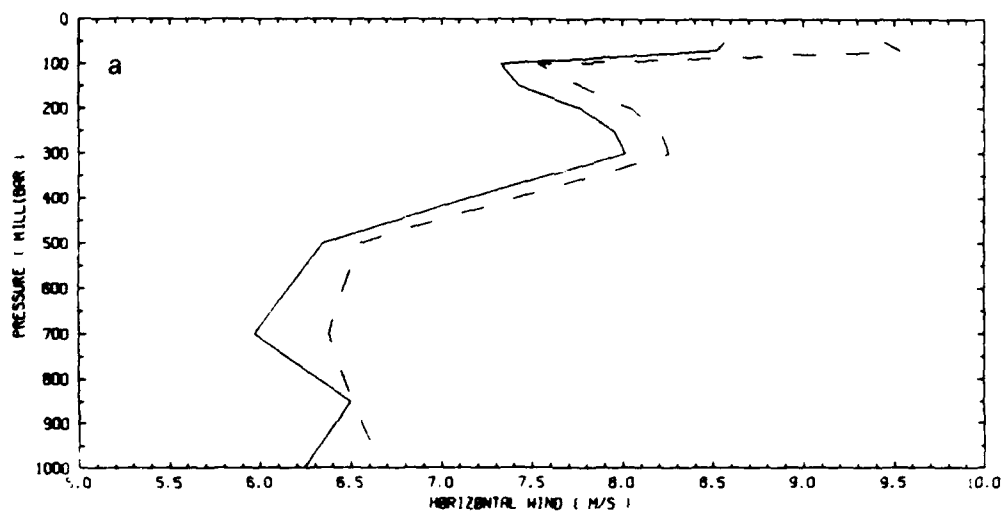


Fig. 9. Time and global rms analysis error for the STATSAT (dashed) and OITEST (solid) assimilation experiments for (a) vector wind error (m/s) and for (b) height error (m).

While it is not possible to determine which enhancements were most significant in this impact from the experiments conducted, we hypothesize that the impact at the uppermost levels is due to the enhanced treatment of satellite temperature profile data.

#### 4. CONCLUSIONS

The changes made to ASAP have enhanced its consistency and operation. Separate studies of the satellite profile data preprocessing and the specification and solution of the normal equations demonstrated that the changes made are desirable. The overall positive impact of the enhancements is small as measured by a one week data assimilation experiment based on simulated data.

Specific conclusions which may be drawn from the experiments described in this report are:

- \* The new satellite temperature profile preprocessor SATLTMP 3.0 is simpler, more consistent with procedures of the data producer, more efficient in terms of computer time and more accurate. SATLTMP 3.0 is based on the fewest assumptions and is designed so that errors made in individual layers tend to cancel.

- \* The new specification and solution of the normal equations is more consistent and results in smaller estimated analysis errors.

- \* The old method of solution, while correct for positive definite systems, fails to properly trap non-positive definite systems. This yields incorrect solutions, some of which are not trapped by later checks. It may be that these solutions introduced some noise into the previous analyses.

- \* The one week impact experiment showed a significant but small improvement due to the enhancements made. It is not possible to determine which enhancements were most significant in this impact from the experiments conducted.

#### 5. ACKNOWLEDGMENTS

This work benefitted from many encouragements from and helpful discussions with D. Norquist (AFGL/LYP).

## 6. REFERENCES

- Bergman, K. H., 1978: Role of observational errors in optimum interpolation analysis. Bull. Am. Meteorol. Soc., 59, 1603-1611.
- Bergman, K. H., 1979: Multivariate analysis of temperature and winds using optimum interpolation. Mon. Weather Rev., 107, 1423-1444.
- Dey, C. H., and L. L. Morone, 1985: Evolution of the National Meteorological Center Global Data Assimilation System: January 1982 - December 1983. Mon. Weather Rev., 113, 304-318.
- Dey, C. H., P. A. Phoebus, R. E. Kistler, A. J. Desmarais, J. T. Tuccillo and B. A. Ballish, 1985a: 1984 Summary of NMC Operation Global Analysis, Office Note 309. National Meteorological Center, Washington, DC, 11 pp.
- Dey, C. H., W. J. Bostelman and C. P. Arnold, 1985b: Design of a WINDSAT observing system simulation experiment. In Proceedings of the NASA Symposium on Global Wind Measurements, Edited by W. E. Baker and R. J. Curran. A. Deepak Publishing, Hampton, VA, pp. 73-80.
- Dongarra, J. J., C. B. Moler, J. R. Bunch and G. W. Stewart, 1979: LINPACK User's Guide, SIAM, Philadelphia.
- Gerlach, A. M., (ed.) 1983: Objective Analysis and Prediction Techniques - 1983. AFGL-TR-83-0333, Air Force Geophysics Laboratory, Hanscom AFB, MA, 219 pp. [NTIS ADA 142441]
- Hoffman, R. N., 1987: Documentation of the AFGL statistical analysis program (ASAP) for the global multivariate analysis of heights and winds. AFGL-TR-87-0155, Air Force Geophysics Laboratory, Hanscom AFB, MA, 92 pp. [NTIS ADA 182600]
- Louis, J.-F., R. N. Hoffman, T. Nehrkorn, T. Baldwin and M. Mickelson, 1987: Baseline observing system experiments using the AFGL global data assimilation system. AFGL-TR-87-0327, Air Force Geophysics Laboratory, Hanscom AFB, MA, 92 pp. [NTIS ADA 188861]
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon. Weather Rev., 109, 701-721.
- Norquist, D. C., 1986: Alternative forms of moisture information in 4-D data assimilation. AFGL-TR-86-0194, Air Force Geophysics Laboratory, Hanscom AFB, MA, 139 pp. [NTIS ADA 179792]
- Norquist, D. C., 1988: Alternative forms of humidity information in global data assimilation. Mon. Weather Rev., 116, 452-471.
- Stobie, J., 1984: Choleski Method for Solving a Series of Linear Equations. TSIN Office Note 84-1, Air Force Global Weather Central, Offutt AFB, NE.

Stobie, J. G., M. D. Lewis, M. A. Langford, J. P. Tuell and E. L. Carr,  
1985: The use of optimum interpolation at AFGWC. In Preprints, Seventh  
Conference on Numerical Weather Prediction, June 17-20, 1985, Montreal,  
P. Q., Canada, AMS, Boston, pp. 43-49.

WMO, 1986: Formats for International Exchange of FGGE Level II Data Sets,  
International Council of Scientific Unions/World Meteorological  
Organization, The World Climate Research Programme, WMO/TD-No. 100, WMO,  
Geneva, 86 pp.

## APPENDIX. PROGRAM DOCUMENTATION.

This appendix documents the new subroutines and modifications to the existing subroutines in the same format as Hoffman (1987). Each page has two page numbers, one is the page number within this appendix (e.g. A-3) and one is the insert page number within Hoffman (1987) (e.g. Rev. 1.10, 21.3). Pages 21-88 of D should be replaced by the following pages. Changes and additions are marked in the left hand margin.

This page intentionally left blank.

### 3. Program Units

The main program ASAP1 and the various subprograms it involves are described here in alphabetical order. Within each description, there are various numbered subsections. The start of each of these blocks is identified by a comment card in the FORTRAN code.

# ANCHOR

Name: ANCHOR (subroutine)

Purpose: Transforms SATEM thickness temperatures into heights. Thicknesses are anchored to  $Z_*$  or first guess height. To convert pressure layer thickness to sigma layer thickness, a piece-wise linear in  $\ln(p) - q$  temperature profile is constructed.

Author: Ross Hoffman, AER, 1988

Documentation: Ross Hoffman, AER, 1988

Referenced by: SATLTMP

References: None

Commons used: None

Arguments:

$q_l^\wedge$	$l = 0, L$	$\ln p_l^\wedge$ , log pressure at pressure layer interface (mb)
$T_l$	$l = 1, L$	pressure layer temperatures (K)
$L$		number of pressure layers
$q_k^\wedge$	$k = 0, K$	$\ln p_k^\wedge$ , log pressure at sigma layer interface (mb)
$Z_*$		surface height (m)
$Z_k^\wedge$	$k = 1, K$	interface heights (m)
$\hat{Z}_k^\wedge$	$k = 1, K$	first guess interface heights (m)
$K$		number of sigma layers
VMISS		missing value
IFAIL		set to 1 if ANCHOR fails, 0 otherwise
$T_l^\wedge$	$l = 0, L$	workspace for pressure level temperatures
$T_k^\wedge$	$k = 1, K$	workspace for sigma layer temperatures
$T_k^\wedge$	$k = 0, K$	workspace for sigma level temperatures

I/O Units:       None

Description:

(0) Parameters

$g$	gravity	9.81 m/s
$R$	gas constant	$287.05 \text{ (m/s)}^2 \text{ K}^{-1}$
$q_a$	ln anchor pressure	ln 400 mb

(1) If any  $T_\ell = \text{VMISS}$  go to (FAIL)ure

(2) Determine  $m$  such that  $|q_m^\wedge - q_a|$  is minimized.  
If  $m$  is not found or  $m < 1$  or  $m > L-1$  go to (FAIL)ure

(3) Determine  $T_m^\wedge = T_m + [(T_{m+1} - T_m)/(q_{m+1}^\wedge - q_{m-1}^\wedge)](q_m^\wedge - q_{m-1}^\wedge)$

(4) For  $\ell = m+1, L$  determine  $T_\ell^\wedge = 2T_\ell - T_{\ell-1}^\wedge$

For  $\ell = m, 1$  determine  $T_{\ell-1}^\wedge = 2T_\ell - T_\ell^\wedge$

(5) Determine  $T_k^\wedge$ . First find layer including  $T_k^\wedge$ , then interpolate as follows. Set  $\ell = 1$ . Then for  $k = 0, K$ . If  $q_0^\wedge \geq q_k^\wedge \geq q_L^\wedge$ , then  
(else  $T_k^\wedge = \text{missing}$ )

Until  $[q_{\ell-1}^\wedge \geq] q_k^\wedge \geq q_\ell^\wedge$  increment  $\ell$

$T_k^\wedge = T_\ell^\wedge + [(T_{\ell-1}^\wedge - T_\ell^\wedge)/(q_{\ell-1}^\wedge - q_\ell^\wedge)](q_k^\wedge - q_\ell^\wedge)$

(6) Begin determination of  $T_k$ . Find pressure layers  $\ell = m, n$  which cover the sigma layer, as follows. Set  $n = 1$ , then for  $k = 1, K$ .

If  $q_0^\wedge \geq q_{k-1}^\wedge$  and  $q_k^\wedge \geq q_L^\wedge$ , then (else  $T_k = \text{missing}$ )

Set  $m = n$ ; until  $[q_{m-1}^\wedge \geq] q_{k-1}^\wedge > q_m^\wedge$  increment  $m$

Set  $n = m$ ; until  $[q_{n-1}^\wedge >] q_k^\wedge \geq q_n^\wedge$  increment  $n$

# ANCHOR

(7) Accumulate layer contribution. Set  $T_k = 0$ ; then for  $l = m, n$

If  $l = m$ , then  $q_B = q_{k-1}^\wedge$ ,  $T_B = T_{k-1}^\wedge$

else  $q_B = q_{l-1}^\wedge$ ,  $T_B = T_{l-1}^\wedge$

If  $l = n$ , then  $q_T = q_k^\wedge$ ,  $T_T = T_k^\wedge$

else  $q_T = q_l^\wedge$ ,  $T_T = T_l^\wedge$

$T_k = T_k + (q_B - q_T)(T_B + T_T)$

(8) Normalize

$T_k = T_k / (2(q_{k-1}^\wedge - q_k^\wedge))$ ; end for k (6).

(9) Calculate  $Z_k^\wedge$ . For  $k = 1, K$ . If  $T_k$  is not missing (else  $Z_k = \text{missing}$ )

Determine anchor

$$Z_a = \begin{cases} Z_* & \text{if } k=1 \\ Z_{k-1}^\wedge & \text{if } Z_{k-1}^\wedge \neq \text{VMISS} \\ Z_{k-1}^\wedge & \text{otherwise.} \end{cases}$$

Integrate hydrostatically

$$Z_k^\wedge = Z_a + (R/g) T_k (q_{k-1}^\wedge - q_k^\wedge)$$

10) Return and error handling

Name: ASAP1 (main program)

Purpose: Performs the statistical analysis procedure following Bergman (1979) and Dey and Morone (1985) as described in Norquist (1982, 1983, 1984, 1986a, 1986b) and Halberstam et al. (1984). The moisture analysis is now separate but moisture residuals are calculated here for the GWE Level II data and stored with their associated identification and location parameters on units 8 and 3. ASAP1 calls several routines to read in the data, interpolate to sigma coordinates, and perform quality control. It then loops over all the analysis grid points, selecting observations and calling ASAP2 to actually perform the analysis. Note then an observation may be an entire RAOB report or a single level AIREP. The RAOB contains many pieces of data, the AIREP only one.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: None

References: ASAP2, CNTOBS, FLAGS, MASTOR1, MASTOR2, MASTOR4, MASTOR6, POINTS, SETFG.

Commons used: AS140, AS1AS2, CCONST, DCONST, RESID, SIGK

Arguments: None

I/O units: 2, 3, 7, 8, 10, INPUT, OUTPUT

Description:

(1) Initialize constants. Print version to OUTPUT.

ASAP1

- (2) Read in namelist containing variables describing date and time and whether this is the first analysis in a sequence.
- (3) Open unit 2, the unpacked GWE Level II data set, described by Norquist (1984, p. 17-33) and check header record date/time against that read via namelist (2).
- (4) Set up the vertical structure in /SIGK/. (See Section 5.4.)
- (5) Read in type 1 data (RAOBS) from file 2 on unit 2 (MASTOR1). In (5) - (8) the call to SETFG initializes the horizontal interpolation procedure.
- (6) Read in type 4 data (SATEMS) from file 5 on unit 2 (MASTOR4).
- (7) Read in type 2 data (AIREPS) from file 3 on unit 2 (MASTOR2).
- (8) Read in type 6a data (SATWINDS) from file 7 on unit 2 (MASTOR6).
- (9) Count valid data read in and display totals (CNTOBS).
- (10) Gross and buddy check data (FLAGS), assign points to each observation (POINTS). These points will be used in the observation selection procedure. Count remaining data and display totals (CNTOBS).
- (11) Store residuals on unit 8 and identification/location parameters on unit 3.
- (12) Beginning of analysis phase [(12) - (27)]. Select tropospheric value of  $k_h$  ( $2.0E-12 \text{ m}^{-2}$ ) given by Dey and Morone (1985). (See (23).) Begin loop over all latitudes in the analysis grid.
- (13) Determine  $\phi_g$ , the latitude of the current grid row. The sine of  $\phi_g$  is read from unit 7 if Guassian latitudes are used (ITLAT = 1), otherwise regularly spaced  $\phi_g$  are used. (Currently ITLAT is hard-wired to 1.)

- (14) If this is the first analysis in a sequence read in the standard estimated prediction errors (EPE) at the mandatory levels from unit 10. In this case the EPE,  $E_p(\hat{Z}_\ell)$ ,  $E_p(\hat{U}_\ell)$ ,  $E_p(\hat{V}_\ell)$  do not depend on longitude.
- (15) Begin loop over all longitudes. Define  $\lambda_g$ .
- (16) If  $|\phi_g| \geq 70$ , convert  $(\lambda_g, \phi_g)$  to  $(x_g, y_g)$  the coordinates in a polar stereographic projection true at the pole. (See Section 5.5.)
- (17) Define the latitude/longitude observation selection box. The box is  $\pm 15^\circ$  in latitude and  $\pm 30^\circ$  in longitude if  $|\phi_g| \leq 60^\circ$ . (Longitude window would be  $\pm 40^\circ$  if  $k_h < 1.8 \text{ E} - 12 \text{ m}^{-2}$ , but this will not occur for current parameter settings.) For  $|\phi_g| > 60^\circ$  all longitudes will be considered. (See (20) below.)
- (18) Begin data selection phase [(18) - (25)]. (See Norquist 1986b, pp. 9-10.) Set  $\mu_{\min}^{zz}$  to 0.1. Start loop on data type, NTY = 1, 2, 3, 4, which correspond to type = 1, 4, 2, 6a. If  $M_{\max}$  (i.e. NSNDS) observations have been selected exit loop. If there are no observations for this type go to end of loop on type (25). Begin loop on observations within this type. This algorithm implies that if  $M_{\max}$  type 1 observations are found with  $\mu^{zz} > \mu_{\min}^{zz}$  other data types are not considered, but all the type 1 observations are considered. Similarly for type 1 and 4 combined, type 1, 4, and 2, combined, etc.
- (19) Check that  $\phi_n$  is within the selection box. If not make use of fact that data are sorted by latitude. If  $\phi_n$  is south of box go to end of observation loop. If  $\phi_n$  is north of box go to end of type loop.
- (20) If  $|\phi_g| \leq 60$ , check that  $\lambda_n$  is within the selection box. If not go to end of observation loop. Note the check when the box straddles Greenwich, i.e. when  $\lambda_{\max} < \lambda_{\min}$ . In this case  $\lambda_{\max} < \lambda_n < \lambda_{\min}$  implies  $\lambda_n$  is not in the box.

- (21) If  $|\phi_g| \geq 70$  then transform observation location  $(\lambda_n, \phi_n)$  to polar stereographic projection coordinates  $(x_n, y_n)$  and calculate  $d_{gn}$  the distance between the grid point and the observation.
- (22) Else if  $|\phi_g| < 70$  then calculate distance with approximate great circle formula (Schlatter, 1975).

$$d_{gn}^2 = a^2 \left[ (\cos(\bar{\phi}) \Delta\lambda)^2 + (\Delta\phi)^2 \right]$$

- (23) Calculate horizontal correlation  $\mu_{gn}^{zz} = \exp(-k_h d_{gn}^2)$ . If the number of observations selected,  $M < M_{max}$  skip observations with  $\mu < \mu_{min}$ . If  $M \geq M_{max}$ , skip observations with  $\mu^* < \mu_{min}^*$ . Here \* indicates that  $\mu$  has been multiplied by the number of points assigned to the observation in (10).
- (24) Insert the current observation in the list of observations for this analysis point stored in descending  $\mu^*$  sort order. If  $M$  was already  $M_{max}$ , drop the last one in the list and reset  $\mu_{min}^*$ . If  $M$  was  $M_{max} - 1$ , set  $\mu_{min}^*$  to the last  $\mu^*$  in the list. The list is maintained in two arrays:  $NAKEY(m) = n$ , which points to the observation arrays and  $MUTTXY(m) = \mu_n^*$ .
- (25) End of data selection loops on observation and type (18).
- (26) Apply statistical interpolation procedure. (ASAP2).
- (27) End of loops over analysis longitudes (15) and latitudes (12).
- (28) End of main module.

Name: ASAP2 (subroutine)

Purpose: Performs update of all variables ( $\hat{Z}, \hat{U}, \hat{V}$ ) at all sigma layers for a single ( $\lambda_g, \phi_g$ ) grid location. The procedure follows that of Dey and Morone (1985) and Bergman (1979) as described by Norquist (1986b, pp. 10-18).

In brief, ASAP2 obtains (a) the first guess values and (b) the expected prediction errors at the analysis point. Then, for each layer, including a subsurface layer, ASAP2 (c) selects observations highly correlated with the grid point variables, (d) calculates the observation location-grid point prediction error correlations for the selected observations (i.e., the r.h.s. of the normal equation), (e) calculates the matrix of observation location-observation location prediction and observational error correlations (the same matrix is used for  $\hat{Z}$ ,  $\hat{U}$  and  $\hat{V}$ ), (f) solves the normal equations for the nondimensional analysis weights and (g) calculates the corrections and expected analysis errors. Finally, (h) z corrections are converted to T corrections, and the first guess is updated and stored along with the corrections and estimated analysis errors.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: CHLSKY, ESOBER, LOWTMP, RECALC

Commons used: AS140, AS1AS2, CCONST, DCONST, RESID, SIGK

Arguments: IGP  $\lambda_g$  index  
JGP  $\phi_g$  index  
IMON month of the analysis

## ASAP2

I/O units: 4, 6, 10, 11, OUTPUT

Description: Subscript and pointer usage in ASAP2 are more complicated than elsewhere:

Symbolic name	Meaning
IN, JN	$i, j$ index piece of data which is used in analysis
NVI, NVJ	$D_T(i)$ data type (1,2,3) = (z,u,v) or (1,2) = (z,(u,v))
NN	$m(i)$ intermediate pointer to $n$ and $l$ once $D_T(i)$ is known
NI, NJ	$n(i) = n(m(i), D_T(i))$ observation index within local array
LI, LJ	$l(i) = l(m(i), D_T(i))$ layer/level index
ZB(NI,LI), etc.	$x(i) = x(l(i), n(i), D_T(i))$ is the value of the $i$ th datum. It is stored in one of three arrays for $z$ , $u$ and $v$ separately, depending on the value of $D_T(i)$ .
PH(NI,LI), etc.	$p(i) = p(l(i), n(i), D_T(i))$ is the value of pressure associated with the $i$ th datum. It is stored in one of two arrays for $z$ and $(u,v)$ separately, depending on the value of $D_T(i)$ .
E	expected standard deviation, e.g.,
$E_p(\hat{z}_k)$	EPE of the first guess
$E_a(z_{kg})$	EAE (becomes $E_{a0}(\hat{z}_k)$ )
$E_{a0}(\hat{z}_k)$	EAE of previous forecast
$E_0(u_{ln})$	EOE

(a) - (1) - (2) Obtain the first guess values

- (1) From unit - 4, read in first guess values ( $\bar{T}_k, \bar{U}_k, \bar{V}_k, \bar{P}_*$ ) for current location. Compute  $\bar{P}_k$  and  $\bar{P}_u$ . Here and below the subscript u denotes the underground layer.
- (2) Calculated  $G_g$ , the coefficient of geostrophy at the grid point.  
(See Norquist 1986b, p. 14 and Section 5.6.)

(b) - (3) - (7) Determine the expected prediction errors (EPE).

- (3) Begin determination of EPE (expected prediction errors). If this is the first analysis in a sequence (IFRST = 1), then use the a priori NMC values interpolated linearly in  $\ln p$  to the sigma layer pressures  $\bar{P}_k$ . Above 50 mb, the 50 mb value is used, below 1000 mb extrapolation is used. Since the NMC values do not depend on  $\lambda_g$  they are read outside the longitude loop in ASAP1(14) and stored in /ASLAS2/. The EPE values  $E_p(\bar{Z}_k), E_p(\bar{U}_k), E_p(\bar{V}_k)$  are stored in variable EPE. Note that  $E_p(\bar{Z}_k)$  is the EPE of Z at  $\sigma_k$  not at  $\sigma_{k+1}^\wedge$  (see (7)).
- (4) Otherwise, if this is not the first analysis in a sequence (IFRST = 0), then:

From unit - 10, read  $E_{a0}(\bar{Z}_k)$ , the expected analysis error of the previous analysis. Interpolate  $E_{\Delta T}(\bar{Z}_k)$  the EPE growth during the interval between analyses from the mandatory layer values of Dey and Morone (1985, Table 3, p. 309). Also interpolate  $E_\omega(\bar{Z}_k)$ , the climatological maximum analysis height errors from Table 2 of Dey and Morone, and  $E_T(\bar{U}_k)$ , the tropical wind EPE from Table 4 of Dey and Morone. In all three cases the interpolation procedure is identical to that used in (3). Next, force  $E_{a0}$  to be in the interval  $(E_{\Delta T}, E_\omega)$ . Then set  $E_p(\bar{Z}_k) = E_{a0}(\bar{Z}_k) + E_{\Delta T}(\bar{Z}_k)$ , and calculate  $E_p(\bar{U}_k) = E_p(\bar{V}_k)$  using equations (2.12) and (B.11) of Dey and Morone (1985) as necessary

$$E_p(\bar{U}_k) = E_p(\bar{V}_k) = (g/|f|)(2k_h)^{1/2} E_p(\bar{Z}_k)$$

$$E_B(\bar{U}_k) = E_B(\bar{V}_k) = \frac{|\phi_E|}{25} \sin\left[\frac{90}{25}|\phi_g|\right] E_p(\bar{U}_k) + \left[1 - \frac{|\phi_E|}{25}\right] E_T(\bar{U}_k)$$

where  $k_h$  is set as in FLAGS(7). The blended values  $E_B$  are used only if  $|\phi_g| \leq 25$ . (Note that typographical errors in (B.11) have been corrected here.)

- (5) Initialize variables, setting all corrections ( $z_{kg}$ ,  $u_{kg}$ ,  $v_{kg}$ ) to zero and all estimated analysis errors (EAE) to the EPE, e.g.,  $E_a(z_{kg}) = E_p(\bar{z}_k)$ .
- (6) Loop over observations previously selected in ASAP1 (24), i.e., consider observations in the list NAKEY(i),  $i=1, \text{NGBOX}$ . Extract data ( $p_k$ ,  $u_k$ ,  $v_k$ ,  $Q(u_k, v_k)$ ,  $p_k^\wedge$ ,  $z_k^\wedge$  and  $q(z_k^\wedge)$ ) from /RESID/ for each of these observations for each  $k$  for which data is present and store in local arrays. Observations with no data present are skipped. Since layers/levels with no data are skipped, the pressures associated with the data are stored. That is,  $u_{ln}$  is the  $u$  residual at the  $l$ th layer which had wind or height data present in the  $n$ th observation, and the pressure of  $u_{ln}$  is  $p_{ln}$ . The number of layer/levels extracted is stored in  $L_n$ . Local arrays of  $dsi_n$ ,  $\phi_n$ ,  $\lambda_n$  are also formed. If  $|\phi_g| \geq 70$ , ( $u, v$ ) are converted to polar stereographic coordinates. If  $N$ , the number of observations extracted, is zero, go to (49).
- (7) Assign the EPE for the residuals,  $E_p(z_{ln}^\wedge)$ ,  $E_p(u_{ln})$  and  $E_p(v_{ln})$  by finding the grid point sigma layer pressure  $\bar{p}_{k_m}$  closest to  $p_{ln}^\wedge$  for  $z$  or  $p_{ln}$  for ( $u, v$ ) and assigning the corresponding grid point EPE to the observation EPE, e.g.,  $E_p(u_{ln}) = E_p(\bar{u}_{k_m})$ .
- (8) Initialize IWFLG. Begin main loop (8-46) over layers,  $k=1, K+1$ . Layer  $K+1$  is the underground layer. Select  $k_h$  as in FLAGS(7).  $\bar{p}_g$  is the pressure at the grid point (i.e.,  $\bar{p}_k$ , or  $\bar{p}_u$  if  $k=K+1$ ).

(c) = (9)-(14) Selects the data to be used at this layer.

- (9) Loop over all observations  $n = 1, N$ . Set  $L = L_n$ .
- (10) Calculate  $d_{ng}$ , the distance from the observation to the grid point. For  $|\phi_g| \geq 70^\circ$ , the polar stereographic projection is used. Calculate  $G_n$ , the coefficient of geostrophy at the observation, and  $\mu_g^{zz}$ , the  $zz$  horizontal correlation. (See Section 5.6 and Dey and Morone (1985, Eq. 2.8).)

- (11) Begin loop over all layer/levels in the observation  $l = 1, L$ . If a  $z$  residual is present, then add  $n$  and  $l$  to list of potential  $z$  observation and levels (variables  $NZT(m)$ ,  $LZT(m)$ ,  $m=1, NTOTZ$ ). Calculate the vertical correlation between  $p_{ln}$  and  $\bar{P}_g$  using Eq. (2.9) of Dey and Morone (1985). Obtain the estimated observation error (EOE) for this residual,  $E(z_{ln})$  (ESOBBER). Define  $\epsilon_m$  as  $E(z_{ln})/E_p(z_{ln})$ . Calculate  $\rho_{mg}^{zz}$ ,  $\rho_{mg}^{zu}$  and  $\rho_{mg}^{zv}$ , the adjusted total correlations between the height residual and the grid point variables. (Note that  $m$  implies a particular  $l$  and  $n$ , while  $g$  implies  $k$ .) These adjusted correlations are the correlations obtained following Dey and Morone (1985, Eqs. (2.7), (A2), (A3), (B2), (B3) as appropriate), divided by  $(1 + \epsilon_m^2)$ . These adjusted correlations are the normalized weights one would obtain for updates made using each residual separately. The sum of the absolute values of these weights is

$$\rho_m^{(1)} = |\rho_{mg}^{zz}| + |\rho_{mg}^{zu}| + |\rho_{mg}^{zv}|$$

For the underground layer (i.e. if  $k = K+1$ ) the exact same selection process is used. This ensures that the lowest layer and the underground layer analyses are consistent.

- (12) Repeat (11) for wind data. If wind data are present, then: Add  $n$  and  $l$  to list of potential  $(u,v)$  observations and layers (variables  $NW(m)$ ,  $LW(m)$ ,  $m=1, NTOTW$ ). Calculate vertical correlation between  $p_{ln}$  and  $\bar{P}_g$ . Obtain  $E_o(u_{ln})$  and define  $\epsilon_m$  as  $E_o(u_{ln})/E_p(u_{ln})$ . (Note that  $E_p(u_{ln})$  is used for  $v$  as well.) Calculate  $\rho_{mg}^{uz}$ ,  $\rho_{mg}^{vz}$ ,  $\rho_{mg}^{uu}$ ,  $\rho_{mg}^{uv}$ ,  $\rho_{mg}^{vu}$  and  $\rho_{mg}^{vv}$ , the adjusted correlations, as in (11), but here using Eqs (A4)-(A9) or (B4)-(B9) of Dey and Morone (1985). Set

$$\rho_m^{(2)} = \left[ |\rho_{mg}^{uz}| + |\rho_{mg}^{vz}| + |\rho_{mg}^{uu}| + |\rho_{mg}^{uv}| + |\rho_{mg}^{vu}| + |\rho_{mg}^{vv}| \right] / 2$$

End of loop (9).

- (13) Determine the NLRHS (currently 10) largest  $|\rho|$  greater than 0.1 from the combined set of  $\rho_m^{(1)}$  and  $\rho_m^{(2)}$ . At this point, a  $(u,v)$  wind is considered a single piece of data. Pointers to the selected residuals

are saved in  $\bar{m}(i)$ ,  $\bar{D}_T(zi)$ ,  $i-1, \bar{I}$  (variables NNRHS(i), NVRHS(i),  $i-1, \text{NKT}$ ). The insert sort algorithm (see ASAP1(24)) puts these pointers in the list in descending  $|\rho|$  order.

- (14) If  $\bar{I}=0$  and  $k \leq K$ , go to (45). Else if  $\bar{I}=0$  and  $k=K+1$ , go to (46).  
Else  $\bar{I}>0$ ; continue with:

(d) - (15) - (20) Calculate the observation-gridpoint prediction error correlations for the selected residuals. At this point, the u and v wind components begin to be treated separately. Note that in (17)-(19) the correlations are not adjusted by dividing by  $(1 + \epsilon_m^2)$  as in (11) and (12). Note that all correlations for the underground layer are calculated.

- (15) Set  $j=0$ . Loop over selected data  $i-1, \bar{I}$ . Extract  $\bar{D}_T(i)$  and  $n(i)$  and  $\ell(i)$  from  $\bar{m}(i)$ .  
(16) Repeat (10), calculating  $d_{ng}$ ,  $G_n$  and  $\mu_{mg}^{zz}$ .  
(17) If  $\bar{D}_T(i) = 1$ , then the current residual is a height residual: Increment  $j$ . Add it to the list of pointers, by setting  $m(j) = \bar{m}(i)$  and  $\bar{D}_T(j) = 1$  (variables NNRHSV, NVRHSV). Repeat (11) but store  $\rho_{ng}^{zz}$ ,  $\rho_{ng}^{zu}$ , and  $\rho_{ng}^{zv}$  as the three r.h.s. of the normal equations (variable RHSV).  
(18) If  $\bar{D}_T(i)=2$ , then the current residual is a wind residual:  
Add the u component to the list of pointers as in (17). Calculate  $\rho_{mg}^{uz}$ ,  $\rho_{mg}^{uu}$ , and  $\rho_{mg}^{uv}$  as in (12). Store these adjusted correlations as the three r.h.s. of the normal equations.  
(19) If  $\bar{D}_T(i)=2$ , repeat (18) for the v component. Note:  $\bar{D}_T(j) = 3$ . This is the point where a wind residual is treated as two pieces of data.  
Calculate  $\rho_{mg}^{vz}$ ,  $\rho_{mg}^{vu}$ , and  $\rho_{mg}^{vv}$  as in (12) but store as the three r.h.s. of the normal equations.  
(20) End of Loop (15). Set  $I=j$ . The r.h.s. and the pointers  $\bar{D}_T(i)$ ,  $m(i)$  are complete.

(e) - (21)-(32) Calculate the matrix of observation-observation correlations.

- (21) Loop on all pieces of data  $i-1, \bar{I}$ . Extract  $\bar{D}_T(i)$ ,  $n(i)$ ,  $\ell(i)$ ,  $p(i)$  and  $E_p(x(i))$ . Obtain  $E_o(x(i))$  (ESOBER) and set  $\epsilon_i = E_o(x(i))/E_p(x(i))$ .

- (22) Calculate variables needed for the distance calculation and the correlation calculation which depend on the  $i$ th piece of data. Calculate  $G_i$ , the coefficient of geostrophy.
- (23) Loop on all pieces of data  $j=i, I$ . If  $j=i$ , go to (31). Otherwise, obtain  $D_T(j)$ ,  $n(j)$ ,  $l(j)$ ,  $p(j)$ ,  $E_p(x(j))$ ,  $E_G(x(j))$  and  $\epsilon_j$  as in (21).
- (24) Calculate the distance  $d_{ij}$ , the coefficient of geostrophy  $G_j$ , and other variables needed for the correlation calculation.
- (25) Calculate variables for the vertical correlation calculation. Calculate the horizontal correlation  $\mu_{ij}^{zz}$ .
- (26) Set horizontal observation error correlation (Norquist, 1986b, pp. 14-16) as

$$\mu_{0ij}^{zz} = \begin{cases} 1 & \text{if same variable and same observation} \\ & (D_T(j) = D_T(i) \text{ and } n(j) = n(i)) \\ & \text{for type 1 or 2} \\ \exp(-k_{h0} d^2) & \text{for satellite heights} \\ 0 & \text{otherwise} \end{cases}$$

Note that satellite cloud drift winds errors are assumed to be uncorrelated horizontally. In the sorted Level II data, each aircraft report is a separate "profile" and therefore assumed to be uncorrelated here.

- (27) Calculate vertical error correlations. Let  $x = \ln^2 \left[ p(i)/p(j) \right]$   
Define

$$\nu_{0ij} = \begin{cases} (1 + k_{p0} x)^{-1} & \text{for heights, type 1 or 2} \\ (1 + k_{p0}^w x)^{-1} & \text{for winds, type 1 or 2} \\ \text{Interpolate linearly} & \text{for satellite heights} \\ \text{in } x \text{ in look-up table} & \text{(Norquist, 1986b, Table 3)} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\nu_{ij} = (1 + k_p x)^{-1}$$

- (28) For  $|\phi_g| \geq 70$ , calculate  $\rho$  and  $\eta$  using formulas from appendix A of Dey and Morone (1985). Each of the 9 cases zz, zu, zv, uz, uu, uv, vz, vu, vv are handled separately here. The statistical model for  $\rho$  is also used for  $\eta$ , only the constants  $k_h$  and  $k_p$  are replaced by  $k_{h0}$ ,  $k_{p0}$  and  $k_{p0}^w$ . Note that the use of the geostrophic model for observation error intercorrelations has no effect since in every case that  $\mu_0 \nu_0$  is non-zero, the geostrophic model reduces to a factor of 1.
- (29) For  $|\phi_g| < 70$ , calculate  $\rho$  and  $\eta$  using formulas from appendix B of Dey and Morone (1985).
- (30) Set matrix entries  $R_{ij} = R_{ji} = \rho_{ij} + \epsilon_i \epsilon_j \eta_{ij}$ . Go to (32).
- (31) Set matrix entry  $R_{ii} = 1 + \epsilon_i^2$ .
- (32) End of loop (23) on j. End of loop (21) on i.
- (f) - (33) - (39) Solve the normal equations
- (32a) Geostrophic decoupling is effected for grid points equatorward of 11 degrees. This is accomplished by setting all height wind correlations to zero. This makes the analysis of height univariate and the analysis of winds bivariate. The correlation matrix becomes block diagonal, corresponding to two separate coexisting linear systems. This approach could be extended to decouple u and v if desired. Note that  $D_T(i)$  equals 1, 2, or 3 if the ith datum is a z, u, or v value respectively.
- (32b) Obtain least square solutions. First transpose matrix of right hand sides. Then call CHLSKY to obtains solutions. CHLSKY iteratively checks the solution, rejecting highly correlated data as necessary.
- (33) Begin loop on data type  $D_T(g) = 1, 3$ .
- (34) Unpack  $A_E$  and  $a'$  values for this variable.
- (35) (null)

- (36) (null)
- (37) (null)
- (38) (null)
- (39) (null)
- (40) Convert  $a'_i$  to  $a_i$  the dimensionalized weights using (Norquist, 1986b, Eq. 11)
 
$$a_i = a'_i E_g/E_p(x(i))$$
 where  $E_g = EPE(k, D_T(g))$  for  $k \leq K$  and  $E_g = E_p(\bar{z}_1)$  for the underground layer.
- (41) (null)
- (42) Calculate C, the correction.  $C = \sum_i x(i)$ .
- (43) Perform gross error check of C against ZER (250 m) or WER (25 m/s) (Norquist, 1986b, p. 17). If C passes the check, go to (44).
- (43A) Set  $C=0$ ,  $A_E=1$ . If C is a wind, set the flag (IWFLG(K)) so that both u and v corrections are set to zero in (48).
- (44) If  $k=K+1$ , set  $Z_{ug}=C$ ; otherwise, set  $E_a(z_{kg})$ ,  $E_a(u_{kg})$  or  $E_a(v_{kg})$  equal to  $A_E^{1/2}$  times  $E_p(\bar{z}_k)$ ,  $E_p(\bar{u}_k)$  or  $E_p(\bar{v}_k)$  and  $z_{kg}$ ,  $u_{kg}$  or  $v_{kg}$  equal to C if  $D_T(g) = 1, 2$  or  $3$ . (See Eq. 2.11 of Bergman, 1979.) End of loop (33) on  $D_T(g)$ . Go to (46).
- (45) Set  $z_{kg} = u_{kg} = v_{kg} = 0$  and associated EAE to the corresponding EPE. (Like(5) but for single k only.)
- (46) End of main loop (8) on layers.
- (47) Convert  $z_{kg}$  to  $t_{kg}$  (LOWTMP) and update first guess  $\hat{T}_k = \hat{T}_k + t_{kg}$ .
- (48) Update  $(\bar{U}_k, \bar{V}_k)$  by adding  $(u_{kg}, v_{kg})$ ; first converting analysis winds back to  $(\lambda, \phi)$  coordinates if  $|\phi_g| \geq 70$ .
- (49) Output  $\hat{T}_k$ ,  $\bar{U}_k$ ,  $\bar{V}_k$  to unit 6 and  $E_a(z_{kg})$ ,  $E_a(u_{kg})$ ,  $E_a(v_{kg})$  along with  $z_{kg}$ ,  $u_{kg}$  and  $v_{kg}$  to unit 11. Note that only  $E_a(z_{kg})$  will be used for the next analysis (see (4)). Return.

## BILINR

Name: BILINR (subroutine)

Purpose: Interpolates a single field bilinearly.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: FG

References: none

Commons used: none

Arguments: A  $f_{ij}$ , field array  
(X,Y) (x,y) location in grid units  
NX first A array dimension  
NY (not used)  
ANS  $f(x,y)$ , interpolated value

I/O units: none

### Description:

Determine the coordinates (i,j) of the grid point in the lower left hand corner of the grid cell enclosing the point (x,y). Determine  $x' = x - i$ ,  $y' = y - j$ , the (x,y) location with respect to (i,j). Then set:

$$f(x,y) = x'y' f_{i+1,j+1} + x'(1-y') f_{i+1,j} + (1-x') y' f_{i,j+1} \\ + (1-x') (1-y') f_{i,j}$$

Name: CALCRES (subroutine)

Purpose: Calculate residuals for single level data.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: MASTOR2, MASTOR6

References: None

Commons used: DCONST, DEL, FGDATA, UASIGMA

Arguments: PO  $\bar{P}_k$

POH  $\hat{P}_k$

KV  $k_v$

I/O units: none

Description:

- (0) Residuals are returned in /DEL/.
- (1) Set t, z residuals to missing.
- (2) Calculate (u,v) residuals. If  $(U_{k_v}, V_{k_v})$  is missing set (u,v) residuals to missing.

## CHLSKY

NAME: CHLSKY (subroutine)

Purpose: Solves the normal (i.e. least squares) equations  $Ax = b$  where  $A$  is  $N \times N$  and  $b$  is  $N \times M$ , using the Cholesky  $R^T R$  decomposition. For the wind and height analysis  $M = 3$ . This version does no pivoting or shifting of data arrays. For ill-conditioned cases highly correlated variables are effectively removed but remain as placeholders. For example if the  $i$ th variable is to be removed then the  $i$ th equation becomes  $x(i) = 0$ . In addition, if the maximum absolute value of  $x$  is too large or if the computed value of  $A_E = 1 - (x, b)$  is not in  $[0, 1]$  the most highly correlated variable is removed.

Variables  $x$  and  $b$  may share storage. Only the upper half of  $A$  is referenced.  $N$  must be smaller than  $MAXN$  and  $M$  must be smaller than  $MAXM$ , so that work arrays are large enough.

Author: R. Hoffman, AER, 1988

Documentation: R. Hoffman, AER, 1988

Referenced by: ASAP2

References: SPOCO, SPOSL (Linpak routines)

Commons used: None

Arguments:  $A$ ,  $b$ ,  $N$ ,  $M$ ,  $NDIM$ ,  $x$ ,  $A_E$  as described above. The leading dimension of  $A$ ,  $b$ , and  $x$  are all  $NDIM$  in the calling program. Note that because of the use of work arrays  $A$  and  $b$  will not be changed.

I/O Units: None

## Description:

- (0) Declare failure criteria parameters and subroutine arguments. Declare work arrays. Currently the maximum system size is 20 and the maximum number of right hand sides is 3. If these sizes are exceeded the program aborts.
- (1) Initialize work arrays by copying A into WA and b into WB.
- (2) Decompose  $A = R^T R$ . Preset RCOND to 0. If A is singular RCOND is unchanged by SPOCO.
- (3) If algorithm failed because of a small or zero RCOND go to (8). (Since pivots are returned along the main diagonal of WA a possibly cheap estimate of whether A is poorly conditioned might be obtained by examining the pivots.)
- (4) Solve for the x's by backsubstitution.
- (5) Calculate the maximum value of x and the values of  $A_E$ .
- (6) If algorithm failed because of improbable value of XMAX or of  $A_E$  go to (8).
- (7) Copy solutions in WB to x and return.
- (8)-(10) Error handling section: In case of an error remove one of the observations which is most highly correlated with another. Then go back to start of routine. This assures us that the each solution makes use of exactly the same data.
- (8) Find IMAX, the index of the largest offdiagonal element of A. Note that large negative correlations are unlikely and are not considered here.

CHLSKY

- (8.1) Write values to file for later processing.
- (9) Set ith equation to  $x_i = 0$  for  $i = \text{IMAX}$ .
- (10) Try again. Go to (1).

Name: CNTOBS (subroutine)

Purpose: Counts and prints out the number of valid residuals classified by level and by type.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: none

Commons used: DCONST, RESID

Arguments:

NUM	N	total number of observations
INUM		index of last observation of <u>types</u> 1, 4, 2, 6
ICLK	- 0	before consistency checks
	- 1	after consistency checks

I/O units: OUTPUT

## Description:

- (1) Initialize: write heading, zero counters.
- (2) Loop over all observations, switching type index (IT) as necessary.
- (3) At each level, increment height counter if  $z_k^A$  is present.
- (4) At each level, increment wind counter if  $(u_k, v_k)$  is present. End of loop (2).
- (5) Print results. Return.

## CQCV

Name: CQCV (subroutine)

Purpose: Translates quality indicators from GWE level II data sets to "generic quality marks" of Hoffman (1987) and section 2.4 of this report. (See Section 5.2. for type 1 and type 2 data.)

Author: Ross Hoffman, AER, 1988

Documentation: R. Hoffman, AER, 1988

Referenced by: MASTOR1, MASTOR2

References: none

Commons used: none

Arguments:

IQZ	z, quality mark/indicator
IQT	T, quality mark/indicator
IQQ	q, quality mark/indicator
IQW	(u,v), quality mark/indicator
IDS1	- (not used)
ALAT	- (not used)
ALON	- (not used)
P	- (not used)

I/O units: none

### Description:

(0) On input IQZ contains the GWE quality indicator, on output it contains the ASAP quality mark. The same holds for IQT, IQQ and IQW. Establish translation tables IQH and IQV and lookup functions ICQCV and JCQCV. The maximum value of the translated quality indicators will be used as the quality indicator. Note IQH(1, 1) = 0 instead of 1 to force translation based on the value of IQV in this case. Similarly IQV(1, 2) = IQV(1, 3) = 0. ICQCV translates IH and IV, the horizontal and vertical quality indicator stored in digits 1 and 2 of the input quality indicator and

returns their maximum. JCQCV retracts this maximum value to the interval [0,99].

- (1) Determine type: standard, TWOS, COLEBA or UNKNOWN.
- (2) Translate each input quality indicator to a quality mark in place and return.

## ESOBER

Name: ESOBER (subroutine)

Purpose: Returns estimated observation error for a given variable for a given dsi at a given pressure in a given month. Values for satellite heights are from Norquist (1986b, pp. 36-37); all other values are from Dey and Morone (1985, Table 5).

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP2

References: none

Commons used: none

Arguments:	IDS	<u>dsi</u>
	P	P, pressure of the observation
	NV	variable index; 1-Z, 2-u, 3-v
	PM	$\bar{P}_l$ , pressure at the mandatory levels
	MANDLVL	L, the number of mandatory levels
	EOV	$E_o$ , estimated observation error
	IMON	month of the analysis (not used)

### Description:

- (1) For type 2 data (AIREPS): for dsi = 21,22, set  $E_o$  = 4.9 m/s; for dsi = 23,24, set  $E_o$  = 6.0 m/s.
- (2) For type 6a data, dsi = 61 (SATWINDS): if  $p > 400$  mb, set  $E_o$  = 4.2 m/s; if  $p \leq 400$  mb, set  $E_o$  = 7.5 m/s.
- (3) For type 1 data, dsi = 13,14,15 (TWOS, dropwindsondes), set  $E_o$  = 2.0 m/s. For dsi = 16 (constant level balloon), set  $E_o$  = 5.0 m/s.

- (4) For all others, interpolate linearly in  $\ln p$  in EOE table which has four categories (columns) for each mandatory level (rows). If  $p < 50$  mb, the 50 mb value is used. If  $p > 1000$  mb, the 1000 mb value is used. The four categories are

Column	Usage	dsi	NV
1	RAOB heights	11	1
2	RAOB winds	11	2,3
3	PIBAL winds	12	2,3
4	SAT heights	41	1

FG

Name: FG (subroutine)

Purpose: Bilinearly interpolates first guess fields to the observation location. Returns results in /FGDATA/ and in arguments ZSTAR, PO, POH.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: MASTOR1, MASTOR2, MASTRO4, MASTOR6

References: BILINR

Commons used: CCONST, DCONST, FGDATA, FGFLDS, SIGK

Arguments: RLAT  $\phi$   
RLON  $\lambda$   
ZSTAR  $Z_*$   
PO  $\bar{P}_k$   
POH  $\bar{P}_k^\wedge$   
IDSI dsi  
IWAT flag for  $\bar{R}$  computation (see (3)).  
NRDLAT number of latitudes which have been read. (Index of latitude currently stored as second latitude in /FGFLDS/. That is,  $\phi_2 = -90 + (\text{NRDLAT} - 1) \Delta\phi$ .)

I/O units: 1, 5

Description:

(0) Note that  $\bar{P}_1^\wedge = \bar{P}_*$ .

(1) Determine fine grid coordinates (x,y).

- (2) If  $y > \text{NRDLAT}$ , values for a new latitude are needed. Shift northern latitude values to southern latitude values in /FGFLDS/. Read in a new northern latitude, incrementing NRDLAT. Go to (2).
- (3) Interpolate  $\bar{T}$ ,  $\bar{U}$ ,  $\bar{V}$ ,  $\bar{R}$  from /FGFLDS/ to /FGDATA/. Relative humidity is interpolated only if data is type 1 (RAOBS) or if data is type 4 (SATEMS) and  $\text{IWAT} \neq 0$ .
- (4) Interpolate  $\bar{P}_*$  from /FGFLDS/. Set  $\bar{P}_k$  and  $\bar{P}_k^\wedge$  from  $\bar{P}_*$  and /SIGK/.
- (5) If not type 6 (i.e. for RAOBS, AIREPS or SATEMS) interpolate  $Z_*$  from /FGFLDS/ and obtain heights by integrating  $\bar{T}_k$  hydrostatically.

## FLAGS

Name: FLAGS (subroutine)

Purpose: Perform gross and buddy checks on the data stored in /RESID/ (see Norquist, 1986b, pp. 5-8). All comparisons are made within boxes approximately  $10^\circ$  square (see Section 5.3). Gross error check limit is  $\pm 3$  forecast error standard deviations. Buddy check follows Bergman's (1978, 1979) procedure in spirit, with the addition of keep flags. Data which fail the checks are set to missing.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: REJECT

Commons used: AS1AS2, CCONST, DCONST, IFLG, RESID, SIGK

Arguments:	IBOX	list of $B_n$
	JDSI	list of $dsi_n$
	QLAT	list of $\phi_n$
	QLON	list of $\lambda_n$
	PTS	- not used
	NWC	- not used
	NUM	N

I/O units: output

### Description:

- (0) The arrays used to store the data flags and indices are confusing. In the buddy check, within a box at a single layer or level, observations of the same type are compared by pairs  $i, j$ , with  $i < j$ . I will use the notation in this documentation that  $F_j^i$  means "i puts a flag on j", since

$i$  is on top of  $j$  in the symbol. Also I use  $^wF$  for wind flags and  $^zF$  for height flags. If the data  $i$  and  $j$  are consistent, then  $F_i^i = F_j^j = 0$ . If they are inconsistent and  $i$  is of better quality, then  $F_j^i = 1$  and  $F_i^j = 0$ ; if they are of equal quality then  $F_j^i = F_i^j = 1$  and if  $j$  is of better quality then  $F_i^j = 1$  and  $F_j^i = 0$ .

In the code the following symbols are used

$N = i$	$IN = i$	$IFWI(IN,INO) = ^wF_i^j$	$IFZTI(IN,INO) = ^zF_i^j$
$NO = j$	$INO = j-i$	$IFWJ(IN,INO) = ^wF_j^i$	$IFZTJ(IN,INO) = ^zF_j^i$

- (1) Prediction error standard deviations  $E_p(\hat{Z}_\ell)$ ,  $E_p(\hat{T}_\ell)$ ,  $E_p(\hat{U}_\ell)$  and  $E_p(\hat{V}_\ell)$  for the mandatory levels in 5 latitude belts are defined from Table 1 of Norquist (1986b). The constants  $a$  and  $b$  for the buddy check (cf. Eq. 7.1 of Bergman (1979)) are set to 3.0 and 1.5, one half of Bergman's values.
- (2) Begin loop over the 18  $10^\circ$  latitude bands.
- (3) Begin loop over the boxes in this band.
- (4) Calculate the box index for the current box and begin loop over all observations in the box.
- (5) If array overflow occurs (current dimensional constraint is 100 observations) output message and go to (6). Otherwise, add current observation to the list of observations in this box.  $NKB(i) = n(i)$ ,  $i=1$ ,  $N_B$  is the pointer to the location in /RESID/ of the  $i$ th observation in the current box. End of loop (4).
- (6) Loop over all observations in the box  $i=1$ ,  $N_B$ , extracting residuals and quality marks from /RESID/ and storing them in local arrays. If  $|\phi_B| \geq 70^\circ$ , winds are converted to polar stereographic winds.

# FLAGS

- (7) Loop over layers selecting the appropriate  $k_h$  for this layer. (See Dey and Morone, 1985, p. 308). If  $\sigma_k \geq 0.125$   $k_h = 2E-12m^{-2}$ , if  $0.125 > \sigma_k \geq 0.070$ , then  $k_h = 1.5E-12m^{-2}$ , and if  $0.070 > \sigma_k$ , then  $k_h = 1.0 E-12m^{-2}$ .
- (8) Initialize all toss flags and keep counters to zero.
- (9) Begin loop over all observations in the current box. Determine indices of the mandatory levels above and below the layer and level pressures of the current data. That is find  $l_b$ ,  $l_a$ ,  $l_b^\wedge$ ,  $l_a^\wedge$  such that

$$\begin{aligned} \bar{P}_{l_b}^\wedge &\geq \bar{P}_* & \sigma_k > \bar{P}_{l_a}^\wedge \\ \bar{P}_{l_b^\wedge}^\wedge &\geq \bar{P}_* & \sigma_{k+1}^\wedge > \bar{P}_{l_a^\wedge}^\wedge \end{aligned}$$

Also determine the latitude belt for looking up the forecast error standard deviations.

- (10) Interpolate  $E(\bar{Z}_p^\wedge)$  to  $\sigma_{k+1}^\wedge$  and  $E(\bar{U}_p)$ ,  $E(\bar{V}_p)$  to  $\sigma_k$  linear in  $\ln p$ . Below 1000 mb extrapolation is used, above 50 mb the value at 50 mb is used.
- (11) Perform gross error check on current observation at current layer/level. Residuals greater than 3 forecast standard deviations are rejected, by changing their values and associated quality marks to missing. REJECT is called to report the rejected data and the reason for rejection. A list of observations in the current box which have failed one or more checks (gross or buddy) is maintained: NCH(m) for  $m-1$ , ICH is the  $m$ th observation which has failed a check. (See (33) and (38).) Note that at statement label 72 NCHT is 1 if a height error has been detected and 2 if a wind error has been detected. End of loop (9) on observations.
- (12) If there is more than one observation start a loop over all observations  $i=1, N_B$ . This loop sets the initial flags. Set  $n(i) = NKB(i)$ ; this is the location of the  $i$ th observation in /RESID/.

- (13) Determine the coordinates of the  $i$ th observation for distance calculation. A polar stereographic projection is used if  $|\phi_B| \geq 70$ .
- (14) Begin loop over  $j = i+1, N_B$ . Set  $n(j) = NKB(j)$ ; this is the location of  $j$ th observation in /RESID/.
- (15) Calculate coordinates of  $j$ th observation as in (13) and  $d_{ij}$ , the distance between the two observations as in ASAP1 (21-22).
- (16) Calculate the horizontal  $zz$  correlation (Dey and Morone, 1985, Eq. 2.8) and the vertical correlation (ibid., Eq. 2.9) using layer pressures. If  $z_i$  or  $z_j$  is missing go to (19). Otherwise calculate the vertical correlation (Eq. 2.9) using level pressures, and the value of  $\rho^{zz}$  of the  $zz$  total correlation (ibid., Eq. 27). Calculate the r.h.s. and l.h.s. of the buddy check criterion for  $z$  (Bergman, 1979, Eq. 7.1): these are  $(a-b\rho_{ij}^{xx})E_p(x)$  and  $|x_i - x_j|$ .
- (17) Determine  $Q_L(z_i)$  and  $Q_L(z_j)$ , the  $z$  quality levels. (See Section 5.2.)
- (18) If the buddy check criterion is violated for  $z$ , then set toss flags:
  - (a) If  $Q_L(z_i) \leq Q_L(z_j)$  then  $zF_j^i = 1$  and
  - (b) If  $Q_L(z_j) \leq Q_L(z_i)$  then  $zF_i^j = 1$ .

If the criterion is not violated and if  $\rho_{ij}^{zz} \geq 0.75$ , then increment the keep counters  $zK_i, zK_j$ .
- (19) If either wind is missing go to (22). Calculate  $\rho_{ij}^{uu}$  and  $\rho_{ij}^{vv}$  using equations from Dey and Morone (1985; either (A6) and (A9) or (B6) and (B9) as appropriate).
- (20) Determine wind quality levels  $Q_L(u_i, v_i), Q_L(u_j, v_j), Q_L(u_i, v_i), Q_L(u_j, v_j)$ .

## FLAGS

- (21) Set flags for winds as in (18). The buddy check criterion is violated if either u or v components violate their separate tests. The keep counters are incremented only if  $\rho_{ij}^{uu}$  and  $\rho_{ij}^{vv}$  are both  $\geq 0.75$ .
- (22) End loop (14) on j. End loop (12) on i.
- (23) Loop over all observations in the box. Remove all toss flags on observations which have 2 or more keep flags, i.e. if ( $^zK_i > 1$ ) then  $^zF_i^j = 0$  for  $j = 1, N_B$  and similarly for the winds.
- (24) Initialize interactive rejection cycle. The minimum number of toss flags which require an observation to be tossed ( $N_T$ ) is set to 2, unless there are only 2 observations in a box in which case it is set to 1.
- (25) Begin rejection cycle. Initialize sums ( $^wS_i, ^zS_i, N_K$ ). Begin loop over all observations in box,  $i=1, N_B$ . (Arrays IQZ and IQW are used here for  $^wS_i$  and  $^zS_i$ .)
- (26) Accumulate the number of toss flags on i for  $j > i$ . (Process  $F_i^j$  for  $j > i$ ).
- (27) Accumulate the number of toss flags on j by i for  $j > i$ . (Process  $F_j^i$  for  $j > i$ ). Note (25-27) are equivalent to creating the sums
- $$S_i = \sum_{j=1}^{N_B} F_i^j \quad \text{for all } i \text{ and for } z \text{ and } (u,v).$$
- (28) If  $^wS_i < N_T$  and  $^zS_i < N_T$  then increment  $N_K$  the number of observations which are assured of being kept. End loop (25) on i.
- (29) If  $N_K = N_B$  then all observations remaining are acceptable: go to (35). Otherwise there are more to be tossed.
- (30) Determine minimum and maximum number of toss flags ( $\min S_i, \max S_i$ ) and the index  $i_{\max}$  which corresponds to the observation with the maximum, for both z and (u,v), separately.

- (31) If  $\max S_i$  is  $\geq 4$  truncate to a multiple of 4 (e.g. 7 becomes 4) for both  $z$  and  $(u,v)$ , separately.
  
- (32) Toss data (by setting the local residuals to missing) for those  $i$  which have  $S_i \geq$  the (truncated)  $\max S_i$ . REJECT is called to report the rejected data and the reason for rejection. Maintain a list of pointers to the observations in the local arrays which have been tossed during this rejection cycle., e.g.  $NZ(m)$ ,  $m=1$ , INTZ contains the pointers to the tossed  $z$  observations.
  
- (33) Update the list of pointers to observations which have failed one or more tests ( $NCH(m)$ ,  $m=1$ , ICH) to reflect changes made in (32) using the  $NZ$  and  $NW$  lists.
  
- (34) Remove all toss flags imposed by observations which have just been tossed on other observations. That is for  $i=1, N_B$  if  $S_i \geq$  (truncated)  $\max (S_i)$  then set  $F_j^i = 0$  for  $j=1, N_B$ , for  $z$  and  $(u,v)$  separately.
  
- (35) Go to (25) to start next iteration unless 10 iterations have already been performed.
  
- (36) Loop on observations in the box  $i=1, N_B$ . Restore  $IQZ$ ,  $IQW$  arrays from /RESID/. These arrays ( $Q(z), Q(u,v)$ ) were used to store  $^zS_i$  and  $^wS_i$ . Since some observations have been tossed, set  $Q(z)$  to missing if the  $z$  residual is missing, and set  $Q(u,v)$  to missing if the  $(u,v)$  residual is missing. Check if any observations still might be tossed. If this occurs, print message.
  
- (37) End of loop (7) on layers.
  
- (38) Move changes back to OBD array, using list  $NCH$ . If  $|\phi_B| \geq 70$  convert winds from polar stereographic back to  $(\lambda, \phi)$  coordinates.
  
- (39) End of loops (2) and (3) on buddy check boxes. Print message and return.

## LOWTMP

Name: LOWTMP (subroutine)

Purpose: Converts analyzed residuals  $z_k$  to temperature residuals  $t_k$  and corrects the first guess value  $\hat{T}_k$ . (See Norquist, 1986b, p. 18, pp. 55-58.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP2

References: none

Commons used: DCONST, LOWT

Arguments:

CORZ	$z_k$ , layer analysis
CORZU	$z_u$ , underground layer analysis
TG	$\hat{T}_k$ , first guess on input, updated analysis on output
SL	scaled pressures $p_k$
SLU	scaled pressure $p_u$ in the layer underground
Note:	all arguments are variables for the grid point being analyzed.

I/O units: none

### Description:

(0) Vertical indexing used in this subroutine is illustrated in the following figure. The first value is used in (1) and (4). The second value is used otherwise.

	top of the atmosphere
$k = K+1, K$	
$k = K, K-1$	
.	
.	
.	
$k = 3, 2$	
$k = 2, 1$	
	surface
$k = 1$	

- (1) Compute T corrections between two layers (i.e., at "interfaces"),

$$t_k^\wedge = (z_k - z_{k-1}) / (\ln p_k - \ln p_{k-1}) \quad k = 2, \dots, N_k + 1$$

These temperatures must be multiplied by  $-g/R$  to become dimensionalized (see (3)).

- (2) Reset index so that  $t_k^\wedge = t_{k+1}^\wedge \quad k = 1, \dots, N_k$
- (3) Apply Flattery algorithm, treating  $t_k^\wedge$  as layer temperatures. Since the sigma structure /SIGK/ doesn't change, the required constants are stored in /LOWT/. First form  $BT^\wedge$ , then  $A^T BT^\wedge$  and finally  $(A A^T)^{-1} A^T BT^\wedge$ . Here  $T^\wedge$  is the vector of  $t_k^\wedge$ . Dimensionalize the result by  $-g/R$  to get T, the layer temperatures.
- (4) Set uppermost layer residual to zero:  $t_{K+1} = 0$ . (Note: layers include subsurface layer.) Update temperatures  $\hat{T}_k = \hat{T}_k + t_{k+1} \quad k = 1, K$ .

## MASTOR1

Name: MASTOR1 (subroutine)

Purpose: Reads type 1 data (RAOBS) from unpacked GWE Level II data set, interpolates to sigma and calculates residuals, which are then stored in /RESID/. (See Norquist, 1986b, pp. 31-33.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: CQCV, FG, PTOSIG

Commons used: DCONST, FGDATA, RESID, UADATA, UASIGMA

Arguments:	IDATE	date	
	ITIME	time	
	NW	-	(not used)
	NUM	N	
	QLAT	list of $\phi_n$	
	QLON	list of $\lambda_n$	
	JDSI	list of $dsi_n$	
	IBOX	list of $B_n$	

I/O Units: 2,OUTPUT

### Description:

- (1) Initialize time and date window as  $\pm 3$  hours about date/time which was input. Set ICHNG = 0; this is a flag used to overwrite a previous observation with a new observation of higher quality. Initialize N, L.
- (2) Read a record from unit 2, using format described by Norquist, (1984, pp. 18-19). On EOF go to (19).

- (3) If not acceptable go to (2). Specifically dsi = 75, cloud data, are skipped. The presence of other non-type 1 data results in an ABEND. Records with missing p, negative z or date/time not in the window are skipped.
- (4) If  $\lambda$ ,  $\phi$  or dsi do not match  $\lambda_N$ ,  $\phi_N$ , or dsi<sub>N</sub>, go to (12). That is if current record is not part of current observation, it is time to complete processing the current observation and begin new one. Note that the test makes use of fact that  $\lambda$ ,  $\phi$  are coded only to hundredths of degree.
- (5) If level dimension (currently 90) is exceeded in /UADATA/, skip this level, go to (2).
- (6) If this is not the first level for the current observation, i.e., if  $l \neq 1$ , go to (8). Set  $\lambda_N$ ,  $\phi_N$ , dsi<sub>N</sub> to  $\lambda$ ,  $\phi$ , dsi. Calculate  $B_n$  (see Section 5.3).
- (7) If  $N > 1$ , then loop over all previous observations searching for a matching  $\lambda_n$ ,  $\phi_n$ . If match is found then: If dsi < dsi<sub>n</sub>, then the new observation is better (see WMO (1978), Appendix A, Table 1). Set ICHNG = 1 to force overwriting observation n. (To do this temporarily set  $N = n$ .) Go to (8). Else if dsi  $\geq$  dsi<sub>n</sub>, then flush current records until new location is found. Go to (3).
- (8) Translate T in Celsius to T in Kelvin,  $T_d$  to R, wind speed and direction to (U,V). If quality indicator is 99 for any variable, set corresponding value to missing (DNN = -999.9). If T is missing, or if dew point or dew point depression is unreasonable set R to missing. Winds at the pole are set to missing.
- (9) Combine quality indicators and translate into quality marks (CQCV).

MASTOR1

- (10) Set all quality marks  $> 1$  to 99, set corresponding values to missing. This keeps only data which have been checked satisfactorily and data which have not been checked.
- (11) Increment L. Go to (2).
- (12) Save  $\lambda$ ,  $\phi$  for the new observation which has been detected. If  $L = 1$  (which can happen only if  $N = 1$ ), go to (6).
- (12a) Eliminate duplicate and/or out of order levels from current observation.
- (13) Calculate first guess values at  $\lambda_N$ ,  $\phi_N$  for  $Z_*$ ,  $\bar{P}_k$ ,  $\bar{T}_k$ ,  $\bar{U}_k$ ,  $\bar{V}_k$  and  $\bar{Z}_k^A$  (FG). Although  $IWAT = 0$ ,  $\bar{R}_k$  is calculated, since this is type 1 data.
- (14) If the observation has only a single level skip it by going to (17a). Interpolate observation to sigma coordinates (PTOSIG). If all sigma coordinate data is missing go to (17a).
- (15) Compute residuals. Make sure that missing data results in missing residuals.
- (16) Store residuals and associated quality marks in /RESID/. (At this point quality marks are either 0 for checked residuals, 1 for unchecked residuals, or 99 for missing residuals.)
- (17) Restore N, ICHNG if overwriting. Increment N for next observation.
- (17a) If EOF go to (19). Otherwise more contents of current record to start of /UADATA/ arrays, since this data is start of the new current observation. This resets  $L = 1$ . Go to (6).
- (18) Position unit 2 at start of next file. Go to (12a).
- (19) Undo (17):  $N = N-1$ . Return.

Name: MASTOR2 (subroutine)

Purpose: Reads type 2 data (AIREPS, etc.) from unpacked GWE Level II data set. From each report (u,v) residuals are calculated and assigned to the analysis layer closest to the reported height. (See Norquist, 1986b, pp. 33-34.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: CALCRES, CQCV, FG

Commons used: DCONST, DEL, FGDATA, RESID, UADATA, UASIGMA

Arguments:	IDATE	date	
	ITIME	time	
	NW	-	(not used)
	NUM	N	
	QLAT	list of $\phi_n$	
	QLON	list of $\lambda_n$	
	JDSI	list of $\underline{dsi}_n$	
	IBOX	list of $B_n$	

I/O Units: 2

#### Description:

- (1) Initialize date/time window as  $\pm 3$  hours about input date/time.  
Initialize N for next observation. Initialize /UASIGMA/ humidities to missing ( $Q_k$ ). Initialize all of /UADATA/ to missing ( $P_\ell$ ,  $T_\ell$ ,  $U_\ell$ ,  $V_\ell$ ,  $Q_\ell$ ,  $Z_\ell$ ). Initialize all of /FGDATA/ to missing ( $\hat{Z}_k$ ,  $\hat{T}_k$ ,  $\hat{U}_k$ ,  $\hat{V}_k$ ,  $\hat{R}_k$ ).

# MASTOR2

- (2) Read a type 2 header record. See Norquist, (1984, pp. 20-21). At EOF go to (23).
- (3) Check if data is acceptable. Non-type 2 data results in an ABEND. Data outside date/time window are skipped by going to (22).
- (4) Set  $\phi_N$ ,  $\lambda_N$ , and  $dsi_N$  to values read. Calculate  $B_N$  (see Section 5.3).
- (5) If current observation is collocated with any previous observation, including those of other data types, the current observation is skipped by going to (22).
- (6) Store Z and T and associated quality marks from the header record, converting degrees Celsius to degrees Kelvin.
- (7) Loop on wind records associated with current header record which was read in (2). (Currently sorted data always have a single wind per header record.) First wind report is preset to missing, then data are skipped if quality mark indicates wind information is missing or if location is at the pole. Otherwise wind speed and direction are converted to (U, V) and the quality mark is saved.
- (8) Skip the optional record if present.
- (9) Translate the quality indicators to quality marks (CQCV).
- (10) Only quality marks of 0 and 1 are retained. If  $Q(Z) > 1$ , skip this observation by going to (2). If  $Q(T) > 1$  or  $Q(U,V) > 1$ , the associated values and quality marks are set to missing.
- (11) Interpolate first guess to  $\lambda_N$ ,  $\phi_N$  (FG). Note that  $\bar{R}$  is not interpolated.
- (12) If  $Z < Z_{*}$  or  $Z > \hat{Z}_k$ , then go to (21). Otherwise determine the index  $k_c$  of the  $\hat{Z}_k$  closest to Z.

- (13) Calculate the pressures associated with  $\hat{Z}_k$  and  $Z$  using formulas for a standard atmosphere as described by Norquist (1986b, Eq. 24 and subsequent discussion). Call these  $P(\hat{Z})$  and  $P(\hat{Z})$ .
- (14) The actual pressure associated with the observation is then taken to be  $p$  defined so that  $p = P_{k+1}^{\hat{c}} - P(\hat{Z}) - P(\hat{Z})$ .  
(Note that  $\hat{Z}_k$  is at pressure  $p_{k+1}^{\hat{c}}$ .)
- (15) Determine the sigma layer  $k_v$  which has pressure closest to  $p$ .
- (16) If  $T$  is present, interpolate first guess linear in  $\ln p$  to  $p$  and then add residual to  $\hat{T}_{k_v}$  to obtain  $T_{k_v}$ . This effectively assigns the residual to layer  $k_v$ . If  $T$  is not present set  $T_{k_v}$  to missing.
- (17) If  $(U,V)$  is present interpolate first guess linear in  $\ln p$  to  $p$  and then add the residual to  $(\hat{U}_{k_v}, \hat{V}_{k_v})$  to obtain  $(U_{k_v}, V_{k_v})$ . This effectively assigns the residual to layer  $k_v$ . If  $(U,V)$  is not present, set  $(u_{k_v}, v_{k_v})$  to missing.
- (18) For all  $k \neq k_v$ , set  $T_k, U_k, V_k$  to missing.
- (19) Compute residuals (CALCRES).
- (20) Store results in /RESID/. Each observation contains only  $\hat{P}_*, u_{k_v}, v_{k_v}$ . Increment  $N$ .
- (21) If only a single wind is associated with the current header go to (2). This will always be the case with sorted data. Otherwise process other wind data roughly following steps (4), (11), (12), (13), (14), (15), (17), (19), (20). Then go to (2).
- (22) Skip wind record(s) and the optional record if it is present, which are associated with current header record.
- (23) On EOF position unit 2 after the EOF mark. Undo last  $N$  increment and return.

# MASTOR4

Name: MASTOR4 (subroutine)

Purpose: Reads in SATTEMS from GWE level II unpacked data set which does not include SATTEMS over land. Data is transformed to sigma coordinate height residuals and stored in /RESID/. (See Norquist 1986b, pp. 34-36.) U.S.A. Special Effort quality controlled and/or corrected satellite temperature profiles (dsi - 41) are preferred to raw reports.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: FG, SATLTMP

Commons used: DCONST, FGDATA, RESID, UADATA, UASIGMA

Arguments:	IDATE	date
	ITIME	time
	NW	- (not used)
	NUM	N
	QLAT	list of $\phi_n$
	QLON	list of $\lambda_n$
	JDSI	list of $\underline{dsi}_n$
	IBOX	list of $B_n$

I/O Units: 2

Description:

- (1) Initialize date/time window as  $\pm 3$  hours about date/time as input. Increment N for next observation. Initialize observations interpolated to sigma as missing for U, V and Q, i.e.,  $U_k = V_k = Q_k = DNN$ . Initialize

all first guess values at observation location to missing; i.e.,

$$\hat{Z}_k^\wedge = \hat{U}_k = \hat{V}_k = \hat{R}_k = \hat{T}_k = DNN.$$

- (2) Read in type 4 data record. This should be the group header record for a number of layer records (see Norquist, 1984, pp. 25-27). On EOF go to (18).
- (3) If not acceptable go to rejection loop. Specifically dsi must be 41 and the date/time must be acceptable. Note: soundings over land are not included in the input data set.
- (4) Set  $\lambda_N$ ,  $\phi_N$ , dsi<sub>N</sub> to values read. Calculate  $B_N$ .
- (5) If collocated with a previous observation, possibly of another type, go to rejection loop (17). However if collocated with an observation of the same type and if the data has been corrected by the U.S.A. Special Effort (ISCR - 5) or found correct or probably correct by the U.S.A. Special Effort (IQCI - 6 or 7) then the new observation is processed. When processing is complete instead of incrementing N we shift the contents of the Nth (new) observation to the collocated observation (see (16)).
- (6) Translate quality indicator (IQCI) to quality mark according to WMO (1986), Appendix A, Table 38). If ISCR (WMO (1986), Appendix A, Table 37) indicates observation was corrected set quality mark to 2. All observations with quality marks greater than 1 are skipped by going to rejection loop (17).
- (7) Begin reading layer data records. See Norquist (1984, pp. 26-27) for details on the four types of layer information. MASTOR4 assumes that thickness and mean temperatures are not both present and that standard and nonstandard layer precipitable water are not both present. Obtain first guess values for  $Z_*$ ,  $\hat{P}_k$ ,  $\hat{P}_k^\wedge$ ,  $\hat{T}_k$ ,  $\hat{U}_k$ ,  $\hat{V}_k$ ,  $\hat{Z}_k^\wedge$ , and if layer precipitable water is present,  $\hat{R}_k$ .

- (8) If no thickness data is present go to (12). Read data into arrays of  $P_\ell$  and  $Z_\ell$  where  $P_\ell$  is the pressure at the top of layer  $\ell$ ,  $Z_\ell$  is the height at  $P_\ell$  relative to  $P_1$ . Thus  $Z_1 = 0$ . The layer temperatures  $T_\ell$  are then calculated hydrostatically (see Section 5.1.2). Go to (10).
- (9) For mean temperature data  $P_\ell$  and  $T_\ell$  are read in. Data associated with missing pressures or with layers of zero or negative mass ( $\Delta p$ ) are skipped. The pressure at the lower boundary should match the previous upper boundary pressure. If it is too large an error exists, if too small, then a layer with missing  $T_\ell$  is defined between the two pressures. Indexing used here is normal:  $T_\ell$  is the mean temperature between  $P_\ell$  and  $P_{\ell+1}$ .
- (10) Interpolate  $T_\ell$  to  $T_k$  (at sigma layers) and  $Z_k^\wedge$  (at sigma levels) using Flattery algorithm and hydrostatic relationship along with some first guess information (SATLTMP). Note dimensional limit of 20 layers in matrix ATA.
- (11) Set quality marks for  $Z_k^\wedge$  and  $T_k$  equal to the quality mark defined in (6) for all levels which have data and to 99 for missing data. The layer by layer quality marks of the Level II data are ignored. Go to (13).
- (12) Thickness data is not present; set all quality marks and all  $T_k$ , and  $Z_k^\wedge$  data values, to missing in /UASIGMA/.
- (13) Skip records corresponding to layer precipitable water data. Set all associated quality marks and data values to missing in /UASIGMA/. If mean temperature data is to be read next, then go to (9).
- (14) Begin computation of residuals. Store  $P_*$  in /RESID/. Code is present here to convert Q to R (see Section 5.1.4). However, Q will always be missing (see (13)).
- (15) Calculate residuals; if data is missing in /UASIGMA/, the quality marks and residuals are set to missing. Wind quality marks and residuals are automatically set to missing. All residuals are stored in /RESID/.

- (16) If new point collocates with a previous point move data there. Otherwise increment N. Go to (2).
- (17) Rejection loop: Read thru all records of the current sounding.  
Go to (2). Unexpected EOF results in an ABEND here.
- (18) EOF handling (see (2)).

## MASTOR6

Name: MASTOR6 (subroutine)

Purpose: Reads in type 6a data (SATWINDS) from unpacked GWE Level II data set. From each report, (u,v) residuals are calculated and assigned to the sigma layer closest to the reported pressure. (See Norquist, 1986b, pp. 41-42.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: CALCRES, FG

Commons used: DCONST, DEL, FGDATA, RESID, UADATA, UASIGMA

Arguments:	IDATE	date
	ITIME	time
	NW	- (not used)
	NUM	N
	QLAT	list of $\phi_n$
	QLON	list of $\lambda_n$
	JDSI	list of $\underline{dsi}_n$
	IBOX	list of $B_n$

I/O Units: 2, OUTPUT

### Description:

- (1) Set up quality mark translation tables ICP and ICW (see (7)). Initialize date/time window as  $\pm 3$  hours about input date/time. Increment N for next observation.
- (2) Initialize /UASIGMA/ humidities to missing ( $\hat{Q}_k$ ). Initialize all /UADATA/ to missing ( $P_\ell, T_\ell, U_\ell, V_\ell, Q_\ell, Z_\ell$ ). Initialize all /FGDATA/ to missing ( $\hat{Z}_k, \hat{T}_k, \hat{U}_k, \hat{V}_k, \hat{R}_k$ ).

- (3) Read a type 6a record. See Norquist (1984, pp. 29-30).
- (4) Check if data is acceptable. Non-type 6a data results in an ABEND. Data outside date/time window are skipped by going to (3).
- (5) Set  $\phi_N$ ,  $\lambda_N$ ,  $dsi_N$  to values read. Calculate  $B_N$  (see Section 5.3).
- (6) If current observation is collocated with any previous observation, including an observation of another data type, the current observation is skipped by going to (3). Note collocation criterion agreement to 0.1 degrees in latitude and longitude since for these data location is coded to 0.1 degrees, but other data are coded to 0.01 degrees. However if collocated with an observation of the same type and if the IQC = 6 or 7 indicating the U.S.A. Special Effort found the new observation correct or probably correct, the new observation is processed. The new point overwrites the old point (see (18)).
- (7) Set quality marks Q(T) and Q(U,V) from quality indicators IQCP and IQC read at (3) using the translation tables ICP and ICW. These tables translate WMO (1986) Appendix A Tables 30 and 36 into the ASAP "generic quality marks" (see Section 5.2).
- (8) If T is missing or  $Q(T) > 1$ , set T and Q(T) to missing. Otherwise convert degrees Celsius to degrees Kelvin.
- (9) If wind speed or direction is missing or  $Q(u,v) > 1$ , set (u,v) and Q(u,v) to missing. Otherwise convert wind speed and direction to (u,v).
- (10) If T and (u,v) are missing skip this observation by going to (3).
- (11) Interpolate first guess to  $\lambda_N$ ,  $\phi_N$  (FG). Note that  $\bar{R}$  is not interpolated.

- (12) If reported pressure  $P$  is not between top and bottom sigma layer pressures,  $(\bar{P}_1, \bar{P}_K)$  then skip the current observation by going to (3). Otherwise determine  $k_v$ , the index of the sigma layer pressure closest to the observation pressure.
- (13) If  $T$  is present, interpolate first guess linear in  $\ln p$  to  $P$  and then add residual to  $\hat{T}_{k_v}$  to obtain  $T_{k_v}$ . This effectively assigns the residual to  $k_v$ . If  $T$  is missing, set  $T_{k_v}$  to missing.
- (14) If  $(u,v)$  is present interpolate first guess linear in  $\ln p$  to  $P$  and then add the residual to  $(\hat{U}_{k_v}, \hat{V}_{k_v})$  to obtain  $(U_{k_v}, V_{k_v})$ . This effectively assigns the residual to  $k_v$ . If  $(u,v)$  is missing, set  $(U_{k_v}, V_{k_v})$  to missing.
- (15) For all  $k \neq k_v$ , set  $T_k, u_k, v_k$  to missing.
- (16) Compute residuals (CALCRES).
- (17) Store results in /RESID/. Each observation contains only  $\bar{P}_*, u_{k_v}, v_{k_v}$ .
- (18) If new point collocates with a previous point move data there (see (6)). Otherwise increment  $N$  and go to (3) unless  $N$  now exceeds 6000, the current array dimension in /RESID/.
- (19) EOF or dimensions exceeded. Close unit 2. Undo last  $N$  increment. Return.

Name: POINTS (subroutine)

Purpose: Calculates the number of data quantity points (PTS) for each observation. Data quality of SATEMS is accounted for. Additional credit is given if z and (u,v) are both present. (See Norquist, 1986b, pp. 8-9.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: ASAP1

References: none

Commons used: DCONST, RESID

Arguments: JDSI dsi<sub>n</sub>  
 PTS points<sub>n</sub>  
 NUM N  
 IMON month index

I/O units: none

#### Description:

- (1) Loop over all observations  $n=1, N$ . Initialize points<sub>n</sub> = 0.
- (2) Set factor to 1.0 unless dsi<sub>n</sub>/10<sup>-4</sup> (i.e. SATEMS) in which case factor is set to the ratio of the observational standard deviations of RAOBS to SATEMS (0.42 in winter and 0.53 in summer).
- (3) Loop over all layers  $k = 1, K$ .  
 If  $z_{kn}^{\wedge}$  is present, add factor to points<sub>n</sub>.  
 If  $(u,v)_{kn}$  is present add factor to points<sub>n</sub>.  
 If  $z_{kn}^{\wedge}$  and  $(u,v)_{kn}$  are present add factor to points<sub>n</sub>.
- (4) End loop (3) on k. End loop (1) on n. Return.

# PTOSIG

Name: PTOSIG (subroutine)

Purpose: Interpolates RAOBS to sigma. Assumes that temperature varies linearly in  $\ln p$  and hence that height varies quadratically in  $\ln p$ . Both  $T_\ell$  and  $Z_\ell$  observations are used to calculate  $T_k$ , the sigma layer temperatures. The  $T_k$  supplemented by first guess heights as needed are used to calculate the  $Z_k^\wedge$ . Winds and  $\ln$  (specific humidity) are interpolated linearly in  $\ln p$ . (See Norquist 1986b, pp. 32-33, 61-63.)

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: MASTOR1

References: none

Commons used: DCONST, FGDATA, UADATA, UASIGMA

Arguments:	NLVL	L
	ZSTAR	$Z_\star$
	P0	$P_k$
	POH	$P_k^\wedge$

I/O units: none

## Description:

- (1) Begin loop over all sigma levels  $k-1, K+1$ . This loop interpolates  $Z_\ell$  and  $T_\ell$  to a preliminary estimate of  $Z_k^\wedge$ . Find  $\ell_i$ , the first  $\ell-1, L$  such that  $P_\ell < P_k^\wedge$ . Go to (3).
- (2) Fall through loop. If  $P_L$  is within 0.1 mb of  $P_{K+1}^\wedge$ , allow extrapolation of  $z$  by setting  $\ell_i = L$ . Otherwise go to (8).

- (3) If  $\ell_i - 1$  go to (8). (No downward extrapolation.)
- (4) If  $(P_{\ell_i-1} - P_\ell) > 300$ , go to (8).
- (5) Find  $\ell_\ell$ , the first level  $\ell$  at or below  $\ell_i-1$  which has both  $Z_\ell$  and  $T_\ell$  data. Find  $\ell_u$ , the first level  $\ell$  at or above  $\ell_i$  which has both  $Z_\ell$  and  $T_\ell$  data. If either can't be found, go to (8).
- (6) If  $(P_{\ell_\ell} - P_{\ell_u}) > 300$  mb, go to (8).
- (7) Interpolate  $Z_\ell$  and  $T_\ell$  quadratically in  $\ln p$  to obtain the preliminary estimate of  $Z_k^\wedge$ . Note that here  $Z_k^\wedge$  is the height at  $\bar{p}_k^\wedge$ , not at  $\bar{p}_{k+1}^\wedge$ . This algorithm is comprised of Eqs. C7, C3, C6, C2 and C1 of Norquist (1986b, pp. 61-62), except that the constants A and B are multiplied by (-1) everywhere. The quality mark  $Q(Z_k^\wedge)$  is set equal to the maximum of the quality marks associated with the four input data,  $T_\ell$  and  $Z_\ell$  at  $\ell_\ell$  and  $\ell_u$ . Go to (9).
- (8) Do not interpolate. Set  $Z_k^\wedge$  to missing.
- (9) End of loop (1) on levels.
- (10) Compute  $T_k$ ,  $k=1, K$  hydrostatically from preliminary estimates of  $Z_k^\wedge$  calculated in loop (1)-(9). Set  $Q(T_k)$  equal to the maximum of  $Q(Z_k^\wedge)$  and  $Q(Z_{k+1}^\wedge)$ . If either of  $Z_k^\wedge$  or  $Z_{k+1}^\wedge$  is missing, then  $T_k$  is missing.
- (11) Calculate final  $Z_k^\wedge$  for  $k=1, K$  hydrostatically from  $T_k$  and  $Z_b$ . Now  $Z_k^\wedge$  is at  $\bar{p}_{k+1}^\wedge$ .  $Z_b$  is an estimate of  $Z$  at the bottom of layer  $k$ . For  $k=1$ ,  $Z_b = Z_*$ ; for  $k>1$ ,  $Z_b = Z_{k-1}^\wedge$  if it is present or  $z_b = Z_{k-1}^\wedge$  if  $Z_{k-1}^\wedge$  is missing. If  $T_k$  is missing, then  $Z_k^\wedge$  is missing. Quality marks are set so that  $Q(Z_k^\wedge) = Q(T_k)$ .
- (12) Begin loop on layers  $k=1, K$ . This loop interpolates  $(u, v)$  and  $\ln q$ . Find  $\ell_i$ , the first  $\ell=1, L$  such that  $P_\ell < \bar{p}_k$ . Go to (15).

# PTOSIG

- (13) Fall through loop. If  $P_L$  is within 0.1 mb of  $\bar{P}_k$ , use the values and quality marks of  $P_L$  for  $U_k$ ,  $V_k$ ,  $Q_k$ ,  $Q(U_k, V_k)$  and  $Q(Q_k)$ . Go to (21). Otherwise,
- (14) Do not interpolate. Set  $(U_k, V_k)$ ,  $Q_k$  to missing. Go to (21).
- (15) If  $\ell_i - 1$  go to (14). If  $(P_{\ell_i - 1} - P_{\ell_i}) > 300$  mb, go to (14).
- (16) Find  $\ell_\ell$ , the first level at or below  $\ell_i - 1$  which has (U,V) data. Find  $\ell_u$  the first level at or above  $\ell_i$  which has (U,V) data. If either can't be found go to (19).
- (17) If  $(P_{\ell_\ell} - P_{\ell_u}) > 300$  mb, go to (19).
- (18) Interpolate winds at  $\ell_\ell$  and  $\ell_u$  linearly in  $\ln p$  to  $\bar{P}_k$  to obtain  $(U_k, V_k)$ . Set  $Q(U_k, V_k)$  to maximum of  $Q(U_\ell, V_\ell)$  at  $\ell_\ell$  and  $\ell_u$ . Go to (20).
- (19) Do not interpolate. Set  $(U_k, V_k)$  to missing.
- (20) Repeat (16)-(19) for relative humidity. (See Mitchell, 1985.)
- (21) End of loop (12) on layers. Return.

Name: REJECT (subroutine)

Purpose: This subroutine writes a record to the REJECT file  
(unit 9).

Author: R. Hoffman, AER, 1987

Documentation: R. Hoffman, AER, 1988

Referenced by: FLAGS

References: None

Commons used: AS140, DCONST, RESID, SIGK

Arguments: IFLAG Rejection indicator.

K k, vertical level of rejected data.

N n, observation index of rejected data.

The input parameter IFLAG is a two digit integer. The first digit indicates if the rejection is caused by the height (1) or wind (2) data. The second digit indicates if the rejection is caused by the buddy check (1) or the gross check (2). E.g. 21 indicates a (u,v) datum rejected by the buddy check. k and n point to the location of the rejected data in the OBD array.

I/O Units: 9

#### Description:

(0) When first called open REJECT file.

(1) Determine pressure associated with rejected data.

(2) For the nth observation write  $\phi$ ,  $\lambda$ , dsi, IFLAG, K,  $p_{k+1}^{\wedge}$ ,  $z_k^{\wedge}$ ,  $Q(z_k^{\wedge})$ ,  $p_k$ ,  $u_k$ ,  $v_k$ ,  $Q(u_k, v_k)$  to the REJECT file on unit 9. Return.

## SATLTMP

Name: SATLTMP (subroutine)

Purpose: Computes sigma layer temperatures and sigma level heights from satellite observed layer temperatures.

Author: Ross Hoffman, AER, 1988

Documentation: R. Hoffman, AER, 1988

Referenced by: MASTOR4

References: ANCHOR

Commons used: DCONST, FGDATA

Arguments:	P	$P^{\wedge}_l$ , level pressures
	T	$T^{\wedge}_l$ , layer temperatures
	ATA	workspace
	NLVL	L
	ZS	$Z^{\wedge}_k$
	TSIG	$T^{\wedge}_k$
	PSIG	$P^{\wedge}_k$
	POH	$P^{\wedge}_k$
	ZSTAR	$Z^{\wedge}_k$

I/O Units: OUTPUT

### Description:

- (0) Print message identifying version when first called.
- (1) Fill output arrays with missing values. Calculate the log of the pressures.
- (2) Determine the first set of contiguous non-missing values (from K to L-1).
- (3) Call ANCHOR to calculate  $T^{\wedge}_k$  and  $Z^{\wedge}_k$ .
- (4) Print diagnostic output when first called. Return.

Name: SETFG (subroutine)

Purpose: Initializes the bilinear interpolation by reading in the southernmost two latitudes of the first guess grid. Since data are sorted by latitude only two latitudes are needed at any one time.

Author: D. Norquist, SASC, 1980 - 1986

Documentation: R. Hoffman, AER, 1986

Referenced by: AGAP1

References: none

Commons used: CCONST, DCONST, FGFLDS

Arguments: none

I/O units: 1, 5

Description:

(1) Rewind units 1, 5.

(2) For latitudes 1 and 2, read  $\bar{T}$ ,  $\bar{U}$ ,  $\bar{V}$ ,  $\bar{R}$ ,  $\bar{P}_*$  for all longitudes and layers from unit 1. Similarly, read  $Z_*$  for all longitudes from unit 5.

Note: File structure is  $\lambda$ ,  $k$ ,  $\phi$ , while /FGFLDS/ stores data by  $\lambda$ ,  $\phi$ ,  $k$ .

This page intentionally left blank.

#### 4. Input/Output Usage.

Program ASAP1 uses units 1-8, 10, 11, INPUT and OUTPUT. Table 4.1 describes these files briefly. Table 4.2 lists the I/O activity in ASAP1.

Table 4.1

Files used by ASAP1. All files are either formatted (F) or unformatted (U).

Unit	Type	Description
1	U	Fine mesh first guess values for interpolation. Contains: [( $\bar{X}_{ijk}$ , $i=1,I$ , $k=1,K$ , $X=(T,U,V)$ ), ( $\bar{R}_{ijk}$ , $i=1,I$ , $k=1,K_r$ ), ( $\bar{P}_{*ij}$ , $i=1,I$ )], $j=1,J$ where $I=360$ , $J=181$ .
2	F	Unpacked GWE Level II data. See Norquist (1984), pp. 17-33.
3	U	Location information for observations. Contains: [ $N$ ], [ $\phi_n$ , $\lambda_n$ , $dsi_n$ , $B_n$ , <u>points<sub>n</sub></u> ], $n=1,N$ .
4	U	Analysis grid first guess values. Contains: [( $\bar{X}_{ijk}$ , $k=1,K$ , $X=(T,U,V)$ ), ( $\bar{R}_{ijk}$ , $k=1,K_r$ ), $\bar{P}_{*ij}$ ], $i=1,I$ , $j=1,J$ where $I=61$ , $J=62$ .
5	U	Fine mesh topography for interpolation. Contains: [ $Z_{*ij}$ , $i=1,I$ ], $j=1,J$ where $I=360$ , $J=181$ .
6	U	Analysis grid values. Contains: [( $\bar{X}_{ijk}$ , $k=1,K$ , $X=(T,U,V)$ ), $i=1,I$ , $j=1,J$ where $I=61$ , $J=62$ .

7	U	The sines of the Gaussian latitudes $[\sin \phi_j]$ , $j=1, J$ where $J=62$ .
8	U	The residuals used by the analysis. Contains: $\{OBD(n, NW), NW=1, 75\}$ , $n=1, N$ .
9	F	Rejected data. One record per datum rejected $[\lambda_n, \phi_n, \frac{dsi}{n}, \text{flag}, k, p_k^{\wedge}, Z_k^{\wedge}, Q(Z_k^{\wedge}), p_k, u_k, v_k, Q(u_k, v_k)]$
10	U	If this is the first analysis in a sequence, this file contains the NMC EPE at the mandatory levels as a function of analysis grid latitude: $\{(E(\bar{X}_{j\ell}^p), \ell=1, L, X = \{Z, U, V\})\}$ , $j=1, J$ , where $J=62$ , $L=12$ . Otherwise, this file contains the EAE for height from the previous analysis: $\{E_{a0}(\bar{Z}_{ijk}^2), k=1, K\}$ , $i=1, I$ , $j=1, J$ , where $I=61$ , $J=62$ .
11	U	The EAE and the analysis corrections: $\{(E_a(\bar{X}_{ijk}^2), k=1, K, X = \{Z, U, V\})\}$ , $\{x_{ijk}, k=1, K, x = \{z, u, v\}\}$ , $i=1, I$ , $j=1, J$ where $I=61$ , $J=62$ .
INPUT	F	Standard input device. Contains namelist DATIMOP.
OUTPUT	F	Standard output device. Contains printed output.

---

Note: In this table the contents of each record is described within square brackets ([ ]).

Table 4.2

I/O activity. All I/O activity is listed here except PRINTS to unit OUTPUT.

Unit	Activity	Occurrences
1	Rewind	SETFG (1)
	Read	SETFG (2)
2	Open	ASAP1 (3)
	Read	ASAP1 (3), (6), (7), (8)
		MASTOR1 (2), (7)
		MASTOR2 (2), (7), (22)
		MASTOR4 (2), (8), (9), (13), (17)
		MASTOR6 (3)
	Rewind	ASAP1 (7)
	Open-Close	ASAP1 (3), (6), (7), (8)
		MASTOR1 (18)
		MASTOR2 (23)
		MASTOR4 (18)
	Close	MASTOR6 (19)
3	Write	ASAP1 (11)
4	Read	ASAP2 (1)
5	Rewind	SETFG (1)
	Read	SETFG (2)
		FG (2)
6	Write	ASAP2 (49)
7	Read	ASAP1 (13)
8	Write	ASAP1 (11)
9	Write	REJECT (2)
10	Read	ASAP1 (14)
		ASAP2 (4)
11	Write	ASAP2 (49)
Input	Read	ASAP1 (2)

This page intentionally left blank.

## 5. Methods

In this section we document certain methods which are used in ASAP1. This section of documentation is independent of the actual FORTRAN code. Some of these methods are used repeatedly in the code.

### 5.1 Conversions

#### 5.1.1 Vector wind to components

Given wind speed  $|V|$  and direction  $\theta$  the eastward and northward wind components are given by

$$u = - |V| \sin \theta$$

$$v = - |V| \cos \theta$$

Note that  $\theta$  is the direction the wind comes from. (See also Norquist 1986b, Eq. 23.)

#### 5.1.2 Heights to temperatures hydrostatically (and vice versa)

The integrated form of the hydrostatic relationship is

$$- (g/R) \Delta Z = \int_{p_l}^{p_u} T d \ln p$$

where  $g$  is the acceleration of gravity,  $R$  is the gas constant for dry air,  $\Delta Z$  is the thickness of the layer ( $Z_u - Z_l$ ) between pressure levels  $p_l$  and  $p_u$  ( $p_l > p_u$ ),  $T$  is the temperature and  $p$  is pressure. Strictly  $T$  is virtual temperature. If  $T$  is linear in  $\ln p$ , then

$$- (g/R) \Delta Z = \bar{T} \Delta \ln p$$

where  $\bar{T}$  is layer mean temperature is given by  $(T_u + T_l)/2$ . A similar equation is given by Norquist, (1986b, Eq. 5).

#### 5.1.3 Layer to level temperatures

A least squares procedure for converting between layer and level temperatures due to Flattery is described in detail by Norquist (1986b, Appendix A, pp 55-58.) As an alternative, if one level temperature is known,

then all other level temperatures may be inferred from a complete set of layer temperatures if some assumption is made about the functional form of  $T(p)$  between the levels. If  $T$  is linear in  $\ln p$  then one obtains the method described by Norquist (1986b, Appendix D, pp 64-65). In brief, if  $T = a + b \ln p$  between  $p_\ell$  and  $p_u$ , then since  $\int \ln x \, dx = x \ln x - x$  we have

$$\bar{T} \int_{p_u}^{p_\ell} dp = \int_{p_u}^{p_\ell} (a + b \ln p) \, dp$$

or

$$(p_\ell - p_u) \bar{T} = p_\ell (a + b (\ln p_\ell - 1)) - p_u (a + b (\ln p_u - 1))$$

$$T_u = a + b \ln p_u$$

$$T_\ell = a + b \ln p_\ell$$

which are three equations in the three unknowns  $a$ ,  $b$  and either  $T_u$  or  $T_\ell$ . E.g., if  $T_u$  is known, then

$$a = T_u - b \ln p_u$$

$$b = (\bar{T} - T_u) (p_\ell - p_u) / (p_u + p_\ell [\ln(p_\ell/p_u) - 1])$$

#### 5.1.4 Humidity

Dew point,  $T_d$ , dew point depression  $T - T_d$ , specific humidity  $q$  and relative humidity  $r$  are all used in the analysis. These are related by the following expressions. (Dutton, 1976, Chap. 8):

$$r = e/e_s$$

$$e_s = e_o \exp \left[ \frac{L_{lv}}{R_v} \left( \frac{1}{T_o} - \frac{1}{T} \right) \right]$$

$$= 6.11 \exp[19.9274 - 5443.3618/T]$$

$$e = e_s(T_D)$$

$$W = (1/\epsilon) e/(p-e); e = \epsilon W p/(1+\epsilon W)$$

$$q = W/(1+W); W = q/(1-q)$$

where

$$\epsilon = (\text{molecular weight air})/(\text{molecular weight water}) = 1.61,$$

$e$  is the vapor pressure,  $e_s$  the saturation pressure, and  $W$  the mixing ratio (i.e. the ratio of the mass of vapor to the mass of dry air). Similar equations are given by Norquist (1986b, Eq. 1-4 and 22) except that volume mixing ratio is used and the definition  $\epsilon$  is reversed in Eq. 1-4.

## 5.2 Generic quality marks

We distinguish (a) quality indicators  $Q_I$ , which are read from the GWE Level II data, (b) generic quality marks,  $Q$ , which are used in the analysis selection procedures and (c) quality levels,  $Q_L$ , which are used in the buddy check procedures.

The quality indicators  $Q_I$  are described by WMO (1978), e.g. Table IV of Appendix A. Ordinarily,  $Q_I = 0$  indicates no quality control check (QCC) was made and  $Q_I = 1, 2, \dots$  indicate decreasing quality was determined during the QCC. However  $Q_I$  has different meanings for different data. In ASAP1 all  $Q_I$  are translated to  $Q$ , which always have the same meaning (see Table 5.1). Note if  $Q(Z_i) < Q(Z_j)$  then  $Z_i$  is better than  $Z_j$ .

Table 5.1 Generic Quality Marks

Quality Mark	Upper Air Quality Indicator	Meaning
0	1	Value found correct during QCC.
1	0	No QCC was made.
2	4	Value found erroneous during QCC, reconstituted value inserted.
3	6	Value found missing during QCC, reconstituted value inserted.
4	7	Value found missing during QCC, new value assigned.
5	5	QCC made, new value probably entered.
6	2	Value found suspect during QCC.
7	8	Value found erroneous during limits checks.
8	3	Value found erroneous during QCC.
9	9	Value missing.

The quality levels  $Q_L$  used in the buddy check are given by Norquist (1986b, pp 7-8).

$$Q_L = \begin{cases} Q+1 \\ 3(Q+3) + 1 \\ 3(Q+3) + 2 \\ 3(Q+3) + 3 \end{cases} \quad \text{if type} = \begin{cases} 1 \\ 2 \\ 4 \\ 6a \end{cases}$$

### 5.3 Buddy Check Boxes

The following procedure determines the buddy check box index  $B_n$ , given the latitude  $\phi$  and longitude  $\lambda$  of observation (Norquist, 1986b, p.5):

- (1) Determine the latitude band  $i = 1, \dots, 18$  from SP  $\rightarrow$  NP in  $10^\circ$  increments by  $i = (\phi + 90)/10 + 1$ . The center of the latitude band is then  $\phi_c = 10i - 95$ .
- (2) Determine the number of boxes for this band as  $n_i = 36 * \cos \phi_c$ .

(3) Determine the longitude index.

$$j = \lambda / \Delta\lambda + 1 \text{ where } \Delta\lambda = 360/n_i.$$

(4) Set  $B_n = 100i + j$ .

Note that, for the CRAY-1,  $\lambda$  and  $\phi$  must be incremented by 0.001 to get reproducible results. This should have no effect since  $\lambda$  and  $\phi$  are coded to hundredths only.

#### 5.4 Vertical grid and interpolation

A sigma coordinate system is used. The interface value  $\sigma_k^\wedge$  are given by  $p_k^\wedge/p_*$  for  $k=1, K+1$ . The layer thicknesses  $\Delta\sigma_k = \sigma_k^\wedge - \sigma_{k+1}^\wedge$  are defined as 0.075, 0.125, 0.150, 0.150, 0.125, 0.075,  $6 \times 0.050$ . In addition an assumed underground layer has  $\Delta\sigma_u = 0.075$ . Indexing starts with the ground  $k=1$  and ends with the top of the atmosphere  $k = K+1$ . The layer  $\sigma_k$  values are given by

$$\sigma_k = [(\sigma_k^{\kappa+1} - \sigma_{k+1}^{\kappa+1}) / ((\kappa+1)\Delta\sigma_k)]^{1/\kappa}$$

where  $\kappa = R/C_p$  (Brenner et al., 1982, Eq. 32; Norquist, 1986b, Fig. 1).

Note that the vertical indexing for height is generally offset by 1. Thus,  $Z_k^\wedge$  is the height at  $\sigma_{k+1}^\wedge$ .

Vertical interpolation is generally linear in  $\ln p$ ; with data  $X_\ell$  and  $X_u$  at  $p_\ell$  and  $p_u$  we have

$$\frac{X - X_\ell}{X_u - X_\ell} = \frac{\ln(p/p_\ell)}{\ln(p_u/p_\ell)}$$

In ASAP1, this relationship is often put in point-slope form using the midpoint of the interval,

$$X = (X_\ell + X_u)/2 + [(X_\ell - X_u)/\ln(p_\ell/p_u)] (\ln p - 1/2 \ln p_\ell p_u)$$

### 5.5 Polar stereographic projection

For gridpoints with  $|\phi_g| \geq 70^\circ$  a polar stereographic projection is used for data selection, buddy checking and performing the analysis. Map factors are used throughout following Bergman (1979) although Dey and Morone (1985) suggest that for a polar cap the map factors are not necessary.

The conversion of  $(\lambda, \phi)$  coordinates to  $(x, y)$  coordinates, in a system in which the positive  $x$  direction points from the pole towards Greenwich, is given by (Bergman, 1979, Eqs. 3.3-3.4)

$$m = 2/(1 + \sin(\pm \phi)).$$

$$x = am \cos \phi \cos \lambda$$

$$y = \pm am \cos \phi \sin \lambda$$

where  $m$  is the map factor,  $a$  is the earth's radius and the  $+$  (or  $-$ ) corresponds to the Northern (or Southern) Hemisphere. The distance  $d_{ij}$  between two points indexed  $i$  and  $j$  is then given by (Bergman, 1979, Eq. 3.5)

$$d_{ij}^2 = \frac{1}{\bar{m}} [(x_i - x_j)^2 + (y_i - y_j)^2]$$

where  $\bar{m} = (m_i + m_j)/2$ .

In the coordinate system chosen for the projection, only the winds at  $90^\circ W$  are unchanged from the  $(\lambda, \phi)$  system. The others must be rotated by (Norquist, 1986b, Eq. 6)

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \sin \lambda & \pm \cos \lambda \\ \mp \cos \lambda & \sin \lambda \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

where  $(u', v')$  is the wind in the stereographic projection and  $(u, v)$  is the wind in the  $(\lambda, \phi)$  system. Again  $+$  (or  $-$ ) corresponds to the Northern (or Southern) hemisphere. The reverse conversion is given by the transpose of the above rotation matrix:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \sin\lambda & \mp\cos\lambda \\ \mp\cos\lambda & \sin\lambda \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix}$$

### 5.6 Coefficient of geostrophy

The coefficient of geostrophy  $G$  decouples the wind and height analysis in the tropics. Bergman (1979, Eq. 3.20) uses

$$\hat{G} = 1 - \exp (|\phi|/20)$$

while Dey and Morone (1985, Eq. B10) use

$$\bar{G} = \begin{cases} 0 & 0 \leq |\phi| \leq 10 \\ [\cos (12|\phi| - 300) + 1]/2 & 10 \leq |\phi| \leq 25 \\ 1 & 25 \leq |\phi| \leq 90 \end{cases}$$

ASAP1 uses (Norquist, 1986b, p14, and Fig. 2)

$$G = \begin{cases} \bar{G}(\phi) \hat{G}(25) & 0 \leq |\phi| \leq 25 \\ \hat{G}(\phi) & 25 \leq |\phi| \leq 90 \end{cases}$$

This page intentionally left blank.

## 6. References

- Bergman, K. H., 1978: Role of observational errors in optimum interpolation analysis. Bull. Am. Meteorol. Soc., 59, 1603-1611.
- Bergman, K. H., 1979: Multivariate analysis of temperature and winds using optimum interpolation. Mon. Weather Rev., 107, 1423-1444.
- Dahlquist, G., and A. Bjorck, 1974: Numerical Methods, Trans. by N. Anderson. Prentice-Hall, 573 pp.
- Dey, C. H., 1983: The NMC optimum interpolation procedure. Presented at NMC Workshop on Vector Processing and the Statistical Analysis of Meteorological Data, Camp Springs, MD.
- Dey, C. H., and L. L. Morone, 1985: Evolution of the National Meteorological Center Global Data Assimilation System: January 1982 - December 1983. Mon. Weather Rev., 113, 304-318.
- Dutton, J. A., 1976: The Ceaseless Wind. McGraw-Hill, NY, 579 pp.
- Halberstam, I., C. Johnson, D. C. Norquist, and S.-L. Tung, 1984: Two methods of global data assimilation. AFGL-TR-84-0260, AFGL, Hanscom AFB, MA., 91 pp. [NTIS ADA155981].
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon. Weather Rev., 109, 701-721.
- McPherson, R. D., K. H. Bergman, R. E. Kistler, G. E. Rasch, and D. S. Gordon, 1979: The NMC operational global data assimilation system. Mon. Weather Rev., 107, 1445-1461.
- Mitchell, K. E., 1985: A comparison of moisture variables in the vertical interpolations of a 4-D data assimilation system. Preprints, Seventh Conference on Numerical Weather Prediction, Montreal. Amer. Meteorol. Soc., pp. 310-317.
- Norquist, D. C., 1982: Development of objective analysis schemes using the method of optimum interpolation. In Objective Analysis and Prediction Techniques, A. M. Gerlach (ed.), AFGL-TR-82-0394, AFGL, Hanscom AFB, MA, pp. 51-55 [NTIS ADA131465].
- Norquist, D. C., 1983: Development and testing of a multivariate global statistical analysis system. In Objective Analysis and Prediction Techniques, A. M. Gerlach (ed.), AFGL-TR-83-0333, AFGL, Hanscom AFB, MA, pp. 10-48 [NTIS ADA142441].
- Norquist, D. C., 1984: Users guide for optimum interpolation method of global data assimilation. AFGL-TR-84-0290, AFGL, Hanscom AFB, MA, 67 pp. [NTIS ADA155929].

- Norquist, D. C., 1986a: Use of supplemental moisture information in the global optimum interpolation analysis of humidity. In Objective Analysis and Prediction Techniques - 1985, edited by A. M. Gerlach. AFGL-TR-86-0002. AFGL, Hanscom AFB, MA, pp. 11-49 [NTIS ADA169746].
- Norquist, D. C., 1986b: Alternative forms of moisture information in 4-D data assimilation. AFGL-TR-86-0194, AFGL, Hanscom AFB, MA, 71 pp. [NTIS ADA179792].
- Stobie, J., 1984: Cholesky Method for Solving a Series of Linear Equations. TSIN Office Note 84-1, Air Force Global Weather Central, Offutt AFB, NE.
- WMO, 1978: Appendix 10, FGGE Data Management Plan. GARP Implementation Operations Plan, Vol. 3, WMO, Geneva.